INNOVACIÓN PEDAGÓGICA, MEDIACIÓN TECNOLÓGICA Y CONTEXTUALIZACIÓN INTERCULTURAL EN EL SISTEMA EDUCATIVO

INNOVACIÓN PEDAGÓGICA, MEDIACIÓN TECNOLÓGICA Y CONTEXTUALIZACIÓN INTERCULTURAL EN EL SISTEMA EDUCATIVO

NOVO MUNDO

Todas nuestras publicaciones son sometidas a revisión doble-ciego
de pares externos (*Peer Review Double Blinded*).

Esta publicación cuenta con licencia Creative Commons Reconocimiento -
NoComercial - SinObraDerivada 3.0 Unported License.



Innovación pedagógica, mediación tecnológica y contextualización intercultural en el sistema educativo

Diseño de cubierta y maquetación: Editorial Novomundo.

© de los textos: los autores

© de la edición: Editorial Novomundo S.A.

ISBN: 978-9942-694-24-9

Formato: Digital PDF (Portable Document Format)

DOI: https://doi.org/10.5281/zenodo.24114966

Guayaquil, 2026

Nota editorial: Las opiniones de esta obra no demuestran necesariamente la opinión de Editorial Novomundo S.A., ni de los editores o coordinadores de la publicación; asimismo, los autores se responsabilizarán de obtener el permiso correspondiente para incluir material publicado en otro lugar.

INNOVACIÓN PEDAGÓGICA, MEDIACIÓN TECNOLÓGICA Y CONTEXTUALIZACIÓN INTERCULTURAL EN EL SISTEMA EDUCATIVO

NOVO MUNDO

ÍNDICE

ÍNDICE DE FIGURAS **29**

ÍNDICE DE TABLAS **35**

INTRODUCCIÓN **50**

CAPÍTULO I

DIVERSIFICACIÓN EPISTÉMICA DE LA FORMACIÓN DOCENTE Y PRÁCTICA PEDAGÓGICA DESCOLONIZADORA EN HUMANIDADES: EL PAPEL MEDIADOR DE LA REFLEXIVIDAD EPISTÉMICA CRÍTICA EN EL PERÚ

Wilder Antonio Vega Alcántara
Universidad Nacional del Centro del Perú **52**

Resumen **53**

Abstract **53**

1.1 INTRODUCCIÓN **54**

1.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **58**

1.3 MATERIALES Y MÉTODOS **59**

1.3.1 Diseño **59**

1.3.2 Instrumentos **59**

1.3.3 Participantes y tamaño de la muestra **60**

1.3.4 Análisis de datos **62**

1.4 RESULTADOS **62**
1.4.1 Modelo de medición **62**
1.4.2 Modelo estructural y contraste de hipótesis **65**
1.4.3 Soporte de hipótesis. **69**
1.4.4 Robustez **70**
1.5 DISCUSIÓN **78**
1.5.1 Limitaciones y líneas futuras **81**
1.6 CONCLUSIONES **82**
1.7 REFERENCIAS BIBLIOGRÁFICAS **83**

CAPÍTULO II

ÉTICA BIOCÉNTRICA, INTEGRACIÓN CURRICULAR ECOPEDAGÓGICA Y CONDUCTA PROAMBIENTAL DOCENTE: UN MODELO DE ECUACIONES ESTRUCTURALES CON MEDIACIÓN PARCIAL EN LA EDUCACIÓN BÁSICA DEL PERÚ

Wilder Antonio Vega Alcántara
Universidad Nacional del Centro del Perú **90**
Resumen **91**
Abstract **91**
2.1 INTRODUCCIÓN **92**
2.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **95**
2.3 MATERIALES Y MÉTODOS **96**
2.3.1 Diseño **96**
2.3.2 Participantes y tamaño muestral **96**
2.3.3 Instrumentos **97**

2.3.4 Análisis 98
2.5 RESULTADOS 99
2.5.1 Evaluación del modelo de medición 99
2.5.2 Modelo estructural y mediación 101
2.5.3 Estabilidad del efecto total 108
2.5 DISCUSIÓN 112
2.6 CONCLUSIONES 115

CAPÍTULO III

GESTIÓN DE LA CALIDAD PEDAGÓGICA, MEDIACIÓN TECNOLÓGICA DOCENTE Y CALIDAD DEL APRENDIZAJE HÍBRIDO: UNA CONTINGENCIA TERRITORIAL EN EL SUR DEL PERÚ

Luis Enrique Rivas Pacherrez
Universidad Nacional de Piura

Ana Consuelo Yovera Espinoza
Universidad Nacional de Piura

Yonny Marvin Navarro Yanayaco
Universidad Nacional de Piura

Noelia Elizabeth Palacios Lopez
Universidad César Vallejo

..... 122
Resumen 123
Abstract 123
3.1 INTRODUCCIÓN 124
3.3 MATERIALES Y MÉTODOS 128
3.3.1 Diseño 128

3.3.2 Participantes y tamaño de la muestra **128**
3.3.3 Instrumentos **129**
3.3.4 Análisis de datos **130**
3.4 RESULTADOS **131**
3.4.1 Modelo de medición **131**
3.4.2 Estadísticos descriptivos y ajuste del modelo **132**
3.4.3 Modelo estructural y mediación **133**
3.4.4 Análisis multigrupo por zona **137**
3.4.5 Robustez **141**
3.5 DISCUSIÓN **145**
3.6 CONCLUSIONES **148**
3.7 REFERENCIAS BIBLIOGRÁFICAS **149**

CAPÍTULO IV

GOBERNANZA INSTITUCIONAL MEDIANTE COMPROMISOS DE GESTIÓN ESCOLAR Y RENDIMIENTO ACADÉMICO: EL PAPEL MEDIADOR DE LA PERMANENCIA ESCOLAR EN LA SECUNDARIA PÚBLICA DEL PERÚ, 2025

Luis Enrique Rivas Pacherrez
Universidad Nacional de Piura

Ana Consuelo Yovera Espinoza
Universidad Nacional de Piura

Yonny Marvin Navarro Yanayaco
Universidad Nacional de Piura

Noelia Elizabeth Palacios Lopez
Universidad César Vallejo **156**

Resumen **157**

Abstract **157**

4.1 INTRODUCCIÓN **158**

4.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **160**

4.3. MATERIALES Y MÉTODOS **162**

4.3.1 Diseño **162**

4.3.2 Instrumentos **162**

4.3.3 Participantes y tamaño de la muestra **163**

4.3.4 Procedimiento **164**

4.3.4 Análisis de datos **165**

4.4 RESULTADOS **165**

4.4.1 Estadísticos descriptivos y correlaciones **165**

4.4.2 Modelo de medida **166**

4.4.3 Ajuste del modelo estructural **170**

4.4.4 Modelo estructural y contraste de hipótesis **170**

4.4.5 Varianza explicada y tamaños del efecto **180**

4.4.6 Sensibilidad al estimador y estabilidad **180**

4.5 DISCUSIÓN **182**

4.6 CONCLUSIONES **185**

4.7 REFERENCIAS BIBLIOGRÁFICAS **186**

CAPÍTULO V

CONOCIMIENTO DEL RAZONAMIENTO ALGEBRAICO TEMPRANO, AUTOEFICACIA DOCENTE E INTEGRACIÓN DEL PENSAMIENTO ALGEBRAICO EN LA ENSEÑANZA DE LAS CIENCIAS: UN MODELO DE ECUACIONES ESTRUCTURALES EN LA FORMACIÓN INICIAL DEL PROFESORADO

Romer Juvenal Javier Quijano
Universidad Nacional Hermilio Valdizán **193**

Resumen **194**

Abstract **194**

5.1 INTRODUCCIÓN **195**

5.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **198**

5.3 MATERIALES Y MÉTODOS **199**

5.3.1 Diseño y participantes **199**

5.3.2 Instrumentos y variables **200**

5.3.3 Análisis de datos **201**

5.4 RESULTADOS **202**

5.4.1 Estadísticos descriptivos y correlaciones **202**

5.4.2 Modelo de medición **203**

5.4.3 Ajuste global del modelo **205**

5.4.4 Modelo estructural y descomposición de efectos **209**

5.4.5 Robustez **210**

5.5 DISCUSIÓN **218**

5-6 CONCLUSIONES **222**

5-7 REFERENCIAS BIBLIOGRÁFICAS **223**

CAPÍTULO VI

PRESIÓN ACADÉMICA, OPORTUNIDAD PERCIBIDA Y CLARIDAD DE LA POLÍTICA DE IA COMO DETERMINANTES DE LA INTENCIÓN DE USO INDEBIDO DE LA IA GENERATIVA EN LA EDUCACIÓN SUPERIOR: EL PAPEL MEDIADOR DE LA DESCONEXIÓN MORAL

Mauricio Acevedo Carrillo
Universidad Privada del Norte **231**

Resumen **232**

Abstract **232**

6.1 INTRODUCCIÓN **233**

6.2.2 Presión académica **236**

6.2.3 Oportunidad percibida **236**

6.2.4 Claridad de la política de IA **237**

6.2.5 La desconexión moral como mecanismo **237**

6.2.6 Heterogeneidad disciplinar **238**

6.3 MATERIALES Y MÉTODOS **238**

6.3.1 Diseño, participantes y tamaño muestral **238**

6.3.2 Medidas y procedimiento **239**

6.3.3 Análisis estadístico **240**

6.4 RESULTADOS **241**

6.4.1 Participantes **241**

6.4.2 Modelo de medida **242**

6.4.3 Modelo estructural **245**

6.4.4 Poder explicativo y predictivo **250**

6.4.5 Mediación 256
6.4.6 Robustez 260
6.4.7 Invarianza de medida y análisis multigrupo 261
6.5 DISCUSIÓN 267
6.5.1 Implicaciones teóricas y prácticas 269
6.5.2 Limitaciones y líneas futuras 270
6.6 CONCLUSIONES 270
6.7 REFERENCIAS BIBLIOGRÁFICAS 271

CAPÍTULO VII

EXPOSICIÓN A PATRONES TEXTILES ANDINOS, VISUALIZACIÓN ESPACIAL Y PENSAMIENTO GEOMÉTRICO EN ESTUDIANTES DE SECUNDARIA DEL PERÚ: UN MODELO DE MEDIACIÓN CON INVARIANZA DE MEDICIÓN

Judith Esther Gavidia Medrano
Universidad Nacional Hermilio Valdizán

Arnulfo Ortega Mallqui
Universidad Nacional Hermilio Valdizán 280

Resumen 281
Abstract 281
7.1 INTRODUCCIÓN 282
7.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 285
7.3 MATERIALES Y MÉTODOS 286
7.3.1 Diseño 286
7.3.2 Participantes y tamaño muestral 286

7.3.3 Instrumentos 287
7.3.4 Análisis de datos 288
7.4 RESULTADOS 289
7.4.1 Modelo de medición, validez discriminante e invarianza 289
7.4.2 Descriptivos y modelo estructural 295
7.4.3 Robustez 300
7.5 DISCUSIÓN 305
7.6 CONCLUSIONES 308

CAPÍTULO VIII

ESTRATEGIAS DIDÁCTICAS DE COMPROMISO INTERACTIVO Y CAMBIO CONCEPTUAL EN MECÁNICA CLÁSICA EN ESTUDIANTES DE SECUNDARIA DE CONTEXTOS PERIURBANOS DEL PERÚ, 2026

Judith Esther Gavidia Medrano
Universidad Nacional Hermilio Valdizán

Arnulfo Ortega Mallqui
Universidad Nacional Hermilio Valdizán 316

Resumen 317
Abstract 317
8.1 INTRODUCCIÓN 318
8.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 321
8.2.1 Compromiso interactivo y cambio conceptual 321
8.2.2 El conflicto cognitivo como puerta de entrada 321
8.2.3 La regulación metacognitiva como consolidación 321
8.2.4 Diferencias por sexo en el mecanismo 322

8.3 MATERIALES Y MÉTODOS **322**
8.3.1 Diseño y participantes **322**
8.3.2 Instrumentos y procedimiento **323**
8.3.3 Estrategia analítica **325**
8.4 RESULTADOS **326**
8.4.1 Características de la muestra **326**
8.4.2 Cambio conceptual entre el pretest y el postest **326**
8.4.3 Modelo de medida **330**
8.4.4 Modelo estructural y comparación con modelos alternativos **331**
8.4.5 Efectos indirectos **335**
8.4.6 Invarianza de medida y análisis multigrupo por sexo **340**
8.5 DISCUSIÓN **344**
8.6 CONCLUSIONES **348**
8.7 REFERENCIAS BIBLIOGRÁFICAS **349**

CAPÍTULO IX

HERRAMIENTAS DIGITALES Y RAZONAMIENTO LÓGICO-MATEMÁTICO EN LA EDUCACIÓN PRIMARIA: EVIDENCIA CUASIEXPERIMENTAL CON INFERENCIA ROBUSTA POR AULA

Lina Lucy Fabian Ojeda
Universidad Nacional de Educación Enrique Guzmán y Valle

Karla Sofia Dongo Ortiz
Universidad Nacional de Educación Enrique Guzmán y Valle

Fiorella Ñahuin Antonio
Universidad Nacional de Educación Enrique Guzmán y Valle 357

Resumen **358**

9.1 INTRODUCCIÓN **359**

9.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **361**

9.2.1 El razonamiento lógico-matemático y sus componentes **361**

9.2.2 Vías por las que las herramientas digitales pueden mejorar el razonamiento **362**

9.2.3 Hipótesis **363**

9.3 MATERIALES Y MÉTODOS **363**

9.3.1 Diseño y participantes **363**

9.3.2 Tamaño de la muestra **363**

9.3.3 Intervención **364**

9.3.4 Instrumento y procedimiento **365**

9.3.5 Análisis estadístico **366**

9.4 RESULTADOS **366**

9.4.1 Participantes y equivalencia inicial **366**

9.4.2 Estructura y fiabilidad del instrumento **367**

9.4.3 Estadísticos descriptivos **369**

9.4.4 Supuestos del análisis de covarianza **373**

9.4.5 Efecto de la intervención **374**

9.4.6 Robustez **379**

9.4.7 Niveles de logro **380**

9.5 DISCUSIÓN **381**

9.6 CONCLUSIONES **388**

9.7 REFERENCIAS 389

CAPÍTULO X

APRENDIZAJE AUTORREGULADO, COMPROMISO ACADÉMICO Y ANSIEDAD MATEMÁTICA EN ESTUDIANTES UNIVERSITARIOS DE PRIMER AÑO DEL PERÚ: UN MODELO DE ECUACIONES ESTRUCTURALES CON MEDIACIÓN

Romer Juvenal Javier Quijano
Universidad Nacional Hermilio Valdizán 399

Resumen 400

Abstract 400

10.1 INTRODUCCIÓN 401

10.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 404

10.3 MATERIALES Y MÉTODOS 405

10.3.1 Diseño y participantes 405

10.3.2 Instrumentos 406

10.3.3 Procedimiento 407

10.3.4 Análisis de datos 408

10.4 RESULTADOS 408

10.4.1 Modelo de medición 408

10.4.2 Estadísticos descriptivos y correlaciones 414

10.4.3 Modelo estructural 415

10.4.4 Análisis de robustez 420

10.5 DISCUSIÓN 424

10.6 CONCLUSIONES 428

10.7 REFERENCIAS BIBLIOGRÁFICAS 429

CAPÍTULO XI

EFECTOS DEL MÉTODO SINGAPUR SOBRE LA RESOLUCIÓN DE PROBLEMAS MATEMÁTICOS EN ESTUDIANTES DE PRIMARIA: UN ESTUDIO CUASIEXPERIMENTAL EN AULAS DE PERÚ, 2025

Ibeth Yusely Garcia Puma
Universidad Nacional de San Antonio Abad del Cusco 436

Resumen 437

Abstract 437

11.1 INTRODUCCIÓN 438

11.1.1 Planteamiento del problema 438

11.1.2 Antecedentes y vacío de conocimiento 439

11.1.3 Pregunta, objetivos y contribución 440

11.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 441

11.2.1 La resolución de problemas como ciclo de cuatro fases 441

11.2.2 Cómo actúa el método Singapur sobre el ciclo 441

11.3 MATERIALES Y MÉTODOS 442

11.3.2 Diseño y participantes 442

11.3.3 Instrumento 443

11.3.4 Intervención y procedimiento 443

11.3.5 Estrategia analítica 444

11.4 RESULTADOS 445

11.4.1 Muestra, equivalencia inicial y atrición 445

11.4.2 Propiedades de medición **446**
11.4.3 Estadísticos descriptivos **447**
11.4.4 Verificación de supuestos **448**
11.4.5 Efecto sobre la resolución total de problemas (H1) **449**
11.4.6 Efectos por fase de resolución de problemas (H2) **454**
11.4.7 Análisis de robustez **458**
11.4.8 Logro del dominio **459**
11.4.9 Verificación de los parámetros de diseño **463**
11.5 DISCUSIÓN **467**
11.5.1 Implicaciones prácticas **469**
11.5.2 Limitaciones y líneas futuras **470**
11.6 CONCLUSIONES **471**
11.7 REFERENCIAS BIBLIOGRÁFICAS **472**

CAPÍTULO XII

LA EXCURSIÓN GEOGRÁFICA COMO ESTRATEGIA DIDÁCTICA VIVENCIAL Y EL RENDIMIENTO ACADÉMICO EN GEOGRAFÍA: EL PAPEL MEDIADOR DE LA COMPRENSIÓN ESPACIAL EN ESTUDIANTES DE SECUNDARIA DEL PERÚ, 2025

Mauricio Acevedo Carrillo
Universidad Privada del Norte **479**
Resumen **480**
Abstract **480**
12.1 INTRODUCCIÓN **481**
12.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **484**

12.2.1 La excursión geográfica como experiencia de aprendizaje **484**
12.2.2 La comprensión espacial como mecanismo **484**
12.2.3 Del pensamiento espacial al rendimiento académico **485**
12.2.4 Hipótesis **485**
12.3 MATERIALES Y MÉTODOS..... **486**
12.3.1 Diseño y participantes **486**
12.3.2 Tamaño de la muestra **486**
12.3.3 Instrumentos y procedimiento **488**
12.3.4 Análisis de datos..... **489**
12.4 RESULTADOS **490**
12.4.1 Estadísticos descriptivos y correlaciones **490**
12.4.2 Modelo de medida..... **494**
12.4.3 Ajuste del modelo estructural **494**
12.4.4 Modelo estructural y contraste de hipótesis..... **495**
12.4.6 Sensibilidad al estimador y estabilidad del efecto..... **503**
12.5 DISCUSIÓN **504**
12.6 CONCLUSIONES..... **511**
12.7 REFERENCIAS BIBLIOGRÁFICAS..... **512**

CAPÍTULO XIII

PEDAGOGÍA DEL MICRORRELATO Y NARRATIVAS TRANSMEDIA EN LA FORMACIÓN DOCENTE EN LENGUA Y LITERATURA: UN MODELO PLS-SEM CON ANÁLISIS MULTIGRUPO EN EL PERÚ, 2025 522

Eugenio Ccanto Mallma

Universidad Nacional Intercultural de la Amazonía **522**
Resumen **523**
Abstract **523**
13.1 INTRODUCCIÓN **524**
13.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **527**
13.3 MATERIALES Y MÉTODOS **529**
13.3.1 Diseño y participantes **529**
13.3.2 Instrumentos **530**
13.3.3 Análisis de datos **531**
13.3 RESULTADOS **532**
13.3.1 Modelo de medida y validez discriminante **532**
13.3.2 Modelo estructural y mediación **538**
13.3.3 Capacidad explicativa y predictiva, y sesgo de método común **542**
13.3.4 Invarianza de medición y análisis multigrupo **547**
13.3.5 Soporte de las hipótesis **548**
13.5 DISCUSIÓN **553**
13.6 CONCLUSIONES **556**
13.7 REFERENCIAS BIBLIOGRÁFICAS **558**

CAPÍTULO XIV

TERTULIAS LITERARIAS DIALÓGICAS Y LECTURA CRÍTICA DE LA NARRATIVA INDIGENISTA EN LA EDUCACIÓN SECUNDARIA PERUANA: UN MODELO DE ECUACIONES ESTRUCTURALES LATENTE DE DOS OLEADAS CON LA EMPATÍA LITERARIA INTERCULTURAL COMO MECANISMO 566

Eugenio Ccanto Mallma
Universidad Nacional Intercultural de la Amazonía 566

Resumen 567

Abstract 567

14.1 INTRODUCCIÓN 568

14.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 571

14.2.1 Diálogo igualitario y lectura crítica 571

14.2.2 La empatía literaria intercultural como mecanismo 571

14.2.3 Equidad del efecto 572

14.3 MATERIALES Y MÉTODOS 572

14.3.1 Diseño y participantes 572

14.3.2 Tamaño de la muestra 572

14.3.3 Intervención e instrumentos 574

14.3.4 Procedimiento y análisis 576

14.4 RESULTADOS 577

14.4.1 Equivalencia inicial y pérdida entre oleadas 577

14.4.2 Medición: fiabilidad, validez e invarianza 582

14.4.3 Descriptivos y diferencias ajustadas 587

14.4.5 Efectos total, directo e indirecto **589**

14.4.6 Robustez y heterogeneidad **594**

14.5 DISCUSIÓN **599**

14.6 CONCLUSIONES **602**

14.7 REFERENCIAS BIBLIOGRÁFICAS **603**

CAPÍTULO XV

ESTRATEGIAS LECTORAS INTERCULTURALES BILINGÜES Y COMPRENSIÓN INFERENCIAL EN AULAS PRIMARIAS QUECHUA–CASTELLANO: UN ESTUDIO CUASIEXPERIMENTAL POR CONGLOMERADOS EN LOS ANDES PERUANOS **614**

Ibeth Yusely Garcia Puma
Universidad Nacional de San Antonio Abad del Cusco **614**

Resumen **615**

Abstract **615**

15.1 INTRODUCCIÓN **616**

15.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS **618**

15.2.1 Inferencia y modelo de la situación **618**

15.2.2 Recursos bilingües y culturales como conocimiento inferencial **619**

15.2.3 Hipótesis **620**

15.3 MATERIALES Y MÉTODOS **620**

15.3.1 Diseño y participantes **620**

15.3.2 Programa de lectura inferencial intercultural bilingüe **621**

15.3.3 Medidas y procedimiento **621**

15.3.4 Análisis de datos **623**

15.4 RESULTADOS **623**

15.4.1 Muestra, desgaste y equivalencia basal **623**

15.4.2 Propiedades psicométricas de la medida **624**

15.4.3 Resultados descriptivos **625**

15.4.4 Supuestos de los modelos y agrupamiento **626**

15.4.5 Efecto del programa sobre la comprensión inferencial (H1) **627**

15.4.6 Cambio entre el pretest y el postest **633**

15.4.7 Tipos de inferencia y especificidad (H2 y H3) **633**

15.4.8 Tamaños del efecto y logro del nivel satisfactorio **634**

15.4.9 Robustez del efecto primario y dosis-respuesta **642**

15.5 DISCUSIÓN **647**

15.5.1 Una ganancia específica que no dependió de la lengua dominante **647**

15.5.2 Magnitud del efecto y diálogo con la literatura previa **647**

15.5.3 Efectos diferenciados entre tipos de inferencia **649**

15.4.4 Translenguaje y pedagogías culturalmente sostenibles **650**

15.5.5 Equidad e implicaciones para la educación intercultural bilingüe **650**

15.5.6 Limitaciones y líneas futuras **651**

15.6 CONCLUSIONES **652**

15.7 REFERENCIAS BIBLIOGRÁFICAS **653**

CAPÍTULO XVI

EDUCACIÓN INTERCULTURAL BILINGÜE Y COMPRENSIÓN LECTORA: ESTUDIO LONGITUDINAL EN COMUNIDADES ESCOLARES RURALES ANDINAS DEL PERÚ, 2024-2026 660

Lina Lucy Fabian Ojeda
Universidad Nacional de Educación Enrique Guzmán y Valle

Karla Sofia Dongo Ortiz
Universidad Nacional de Educación Enrique Guzmán y Valle

Fiorella Ñahuin Antonio
Universidad Nacional de Educación Enrique Guzmán y Valle 660

Resumen 661

Abstract 662

16.1 INTRODUCCIÓN 662

16.1.2 Planteamiento del problema 662

16.1.3 Antecedentes y vacío de conocimiento 663

16.1.4 Pregunta, objetivos y contribución 664

16.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 665

16.2.1 Interdependencia y transferencia de la comprensión lectora 665

16.2.2 La implementación de la EIB como contexto de aprendizaje 666

16.2.3 Hipótesis 666

16.3 MATERIALES Y MÉTODOS 667

16.3.1 Diseño 667

16.3.2 Participantes y tamaño de la muestra 668

16.3.3 Instrumentos y procedimiento **669**

16.3.4 Estrategia analítica **670**

16.4 RESULTADOS **671**

16.4.1 Muestra, desgaste y estructura anidada **671**

16.4.2 Estadísticos descriptivos, propiedades de la medida e invarianza **672**

16.4.3 Trayectorias de crecimiento **680**

16.4.4 Implementación de la EIB, nivel inicial y crecimiento (H1, H2 y H5) **680**

16.4.5 Dinámica intrapersona entre lenguas (H3 y H4) **687**

16.4.6 Análisis de robustez **692**

16.5 DISCUSIÓN **697**

16.6 CONCLUSIONES **701**

16.7 REFERENCIAS BIBLIOGRÁFICAS **702**

CAPÍTULO XVII

RETROALIMENTACIÓN AUTOMATIZADA CON INTELIGENCIA ARTIFICIAL GENERATIVA Y PRECISIÓN SINTÁCTICA EN LA ESCRITURA EN INGLÉS COMO LENGUA EXTRANJERA: ENSAYO ALEATORIZADO DE TRES BRAZOS EN UNIVERSIDADES PERUANAS 710

Diego Alejo Morales Aguero
Universidad Nacional Santiago Antúnez de Mayolo 710

Resumen **711**

Abstract **711**

17.1 INTRODUCCIÓN **712**

17.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS 715
17.2.1 La precisión sintáctica y su medición 715
17.2.2 Retroalimentación correctiva docente, automatizada y generativa 715
17.2.3 El compromiso con la retroalimentación como mecanismo 716
17.2.4 Hipótesis 716
17.3 MATERIALES Y MÉTODOS 717
17.4 RESULTADOS 719
17.4.1 Participantes y equivalencia inicial 719
17.4.2 Fiabilidad y estructura de la medida 720
17.4.3 Estadísticos descriptivos 721
17.4.4 Invarianza longitudinal de la medida 722
17.4.5 Efecto de la retroalimentación 723
17.4.6 Contrastes planificados 724
17.4.7 Categorías de error 729
17.4.8 Mecanismo: el compromiso con la retroalimentación 730
17.4.9 Moderación por la competencia inicial 734
17.4.10 Robustez 738
17.5 DISCUSIÓN 741
17.5.1 Una persistencia que no sigue a la calidad de cada corrección 741
17.5.2 Magnitud y comparación entre fuentes 741
17.5.3 El compromiso como mecanismo 742
17.5.4 Qué errores cambian y para quién 743
17.5.5 Implicaciones para el contexto peruano 744

17.5.6 Limitaciones y líneas futuras.....744
17.6 CONCLUSIONES.....745
17.7 REFERENCIAS BIBLIOGRÁFICAS.....746

CAPÍTULO XVIII

ENSEÑANZA DEL INGLÉS COMO TERCERA LENGUA EN ESCUELAS INTERCULTURALES BILINGÜES RURALES ANDINAS DEL PERÚ: PRÁCTICAS CULTURALMENTE PERTINENTES, BARRERAS CONTEXTUALES, AUTOEFICACIA Y DESEMPEÑO EN UN MODELO DE ECUACIONES ESTRUCTURALES MULTIGRUPO, 2025 755

Diego Alejo Morales Aguero
Universidad Nacional Santiago Antúnez de Mayolo**.....755**

Resumen 756

Abstract 756

18.1 INTRODUCCIÓN.....757

18.1.1 Planteamiento del problema 757

18.1.2 Antecedentes y vacío de conocimiento.....758

18.1.3 Pregunta, objetivos y aporte 759

18.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS.....760

18.2.1 La autoeficacia como mecanismo del aprendizaje de lenguas adicionales.....760

18.2.2 Prácticas didácticas culturalmente pertinentes 761

18.2.3 Barreras contextuales.....761

18.2.4 Generalización entre comunidades de lengua originaria 762

18.3 MATERIALES Y MÉTODOS.....762

18.3.1 Diseño y participantes 762

18.3.2 Instrumentos **764**
18.3.3 Análisis de datos **764**
18.4 RESULTADOS **766**
18.4.1 Características de la muestra **766**
18.4.2 Modelo de medición **768**
18.4.3 Ajuste de los modelos **769**
18.4.4 Trayectorias estructurales (H1–H3) **769**
18.4.2 Modelo de medición **771**
18.4.5 Efectos indirectos (H4–H5) **775**
18.4.6 Invarianza entre comunidades de lengua originaria (H6) **779**
18.5 DISCUSIÓN **787**
18.5.2 Implicaciones prácticas **789**
18.5.3 Limitaciones y líneas futuras **790**
18.6 CONCLUSIONES **791**

ÍNDICE DE FIGURAS

Figura 1-1
Modelo de medición: cargas factoriales, validez discriminante y ajuste global. **67**

Figura 1-2
Modelo de ecuaciones estructurales completo y descomposición del efecto de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora. **72**

Figura 1-3
Distribución de las puntuaciones compuestas y regresión parcial de la práctica pedagógica descolonizadora **75**

Figura 1-4
Potencia a priori por análisis y estabilidad del efecto directo **78**

Figura 2-1
Modelo de medición: cargas estandarizadas, matriz de validez discriminante y criterio de Fornell-Larcker **103**

Figura 2-2
Modelo de medición: cargas estandarizadas, matriz de validez discriminante y criterio de Fornell-Larcker. **107**

Figura 2-3
Potencia a priori, ajuste global comparado y estabilidad del efecto total **110**

Figura 3-1
Modelo de medición, validez discriminante y ajuste de los modelos competidores **135**

Figura 3-2
Mecanismo estructural, descomposición de efectos y robustez del efecto de la gestión **139**

Figura 3-3
Contingencia territorial del efecto de la gestión de la calidad pedagógica **143**

Figura 4-1
Distribución y relaciones observadas de los constructos **168**

Figura 4-2
Validez discriminante, ajuste comparado de los modelos y potencia del cálculo a priori del tamaño de la muestra **172**

Figura 4-3
Modelo estructural, descomposición del efecto y sensibilidad al estimador **177**

Figura 5-1
Evidencia de validez del modelo de medición y de ajuste global **207**

Figura 5-2
Modelo covariante estructural y de medición para la descomposición del efecto del conocimiento del razonamiento algebraico temprano en la integración del pensamiento algebraico dentro de la enseñanza de las ciencias **212**

Figura 5-3
Relación bivariada, varianza explicada por bloques y estabilidad del coeficiente del conocimiento del razonamiento algebraico temprano **216**

Figura 6-1
Modelo de medida de la presión académica, la oportunidad percibida, la claridad de la política de IA, la desconexión moral y la intención de uso indebido. **247**

Figura 6-2
Modelo estructural, tamaños del efecto y distribuciones bootstrap conjuntas de los efectos indirectos específicos sobre la intención de uso indebido **252**

Figura 6-3
Capacidad predictiva fuera de muestra y varianza explicada del modelo **258**

Figura 6-4
Invarianza de medida de los compuestos y análisis multigrupo e ntre estudiantes de carreras STEM y no STEM. **265**

Figura 7-1
Modelo de medida y estructural de la exposición a patrones textiles andinos, la visualización espacial y el pensamiento geométrico **293**

Figura 7-2
Descomposición del efecto de la exposición a patrones textiles andinos sobre el pensamiento geométrico **298**

Figura 7-3
Robustez del efecto de la exposición, invarianza de medición y ajuste comparado del modelo **303**

Figura 8-1
Cambio conceptual entre el pretest y el postest en mecánica newtoniana **329**

Figura 8-2
Modelo de medida y modelo estructural de la mediación serial del cambio conceptual **334**

Figura 8-3
Descomposición del efecto de las estrategias de compromiso interactivo sobre la comprensión conceptual newtoniana **339**

Figura 8-4
Invarianza de medida y diferencias por sexo en el modelo estructural **343**

Figura 9-1
Estructura factorial de la prueba de razonamiento lógico-matemático en el pretest y el postest **371**

Figura 9-2
Efecto de la intervención con herramientas digitales sobre el razonamiento lógico-matemático **377**

Figura 9-3
Robustez del efecto, variación entre aulas y transiciones entre niveles de logro **383**

Figura 10-1
Evidencia de validez del modelo de medición **412**

Figura 10-2
Modelo estructural estimado, descomposición del efecto e índices de ajuste **418**

Figura 10-2
Relaciones observadas entre los constructos y coeficientes de la regresión jerárquica de la ansiedad matemática **422**

Figura 11-1
Geometría del ajuste, robustez de la especificación y contrastes dentro de cada escuela para la resolución total de problema **452**

Figura 11-2
Cambio pretest–postest, efectos ajustados y distribuciones del postest por fase de resolución de problemas **456**

Figura 11-3
Distribuciones de las puntuaciones, probabilidad de dominio y proporción de estudiantes con dominio por aula **461**

Figura 11-4
Potencia planificada y realizada, parámetros de diseño y diferencias por problema y fase **465**

Figura 12-1
Distribución y asociación observadas de la excursión geográfica vivencial, la comprensión espacial y el rendimiento académico en geografía **492**

Figura 12-2
Validez del modelo de medida y ajuste de los modelos competidores **497**

Figura 12-3
Modelo estructural y descomposición del efecto de la excursión geográfica vivencial sobre el rendimiento académico en geografía **501**

Figura 12-4
Potencia a priori, estabilidad del efecto y varianza explicada por ruta **506**

Figura 13-1
Arquitectura del modelo de medida de los compuestos **535**

Figura 13-2
Modelo estructural estimado y descomposición del efecto del diseño pedagógico del microrrelato sobre la competencia didáctica literaria **540**

Figura 13-3
Tamaño del efecto, capacidad explicativa y predictiva, y sesgo de método común **544**

Figura 13-4
Heterogeneidad entre tipos de institución formadora e invarianza de medición de los compuestos **550**

Figura 14-1
Diseño, flujo de participantes y equivalencia inicial de las condiciones **580**

Figura 14-2
Fiabilidad, invarianza de medida y validez discriminante de las escalas **585**

Figura 14-3
Modelo estructural latente y descomposición del efecto de las tertulias sobre la lectura crítica **592**

Figura 14-4
Robustez y heterogeneidad del efecto de las tertulias **596**

Figura 15-1
Efecto del programa en la comprensión inferencial (N = 990 alumnos en 62 aulas). **630**

Figura 15-2
Efectos del programa por tipo de inferencia y por ítem (N = 990; 62 aulas) **637**

Figura 15-3
Distribución de las puntuaciones del postest y logro del nivel satisfactorio (N = 990; 62 aulas) **640**

Figura 15-4
Robustez del efecto del programa, fidelidad de implementación y agrupamiento (N = 990; 62 aulas) **644**

Figura 16-1
Trayectorias observadas de la comprensión lectora y propiedades de la medida **677**

Figura 16-2
Implementación de la educación intercultural bilingüe y trayectorias de comprensión lectora **684**

Figura 16-3
Dinámica intrapersona entre la comprensión lectora en lengua originaria y en castellano **690**

Figura 16-4
Potencia del diseño, retención y estructura anidada de la muestra **694**

Figura 17-1
Precisión sintáctica por condición y medición **727**

Figura 17-2
Tasas de error por categoría gramatical **732**

Figura 17-3
Medida de precisión y mecanismo de compromiso con la retroalimentación **736**

Figura 17-4
Moderación por la competencia inicial y robustez del efecto primario **740**

Figura 18-1
Modelo estructural estimado, correlaciones entre indicadores y validez discriminante **773**

Figura 18-2
Distribuciones bootstrap, descomposición de efectos y plano de las trayectorias de mediación 777

Figura 18-3
Invarianza de la medición, trayectorias estructurales por grupo y medias latentes entre comunidades de lengua originaria **782**

Figura 18-4
Relación entre autoeficacia y desempeño por institución, barreras contextuales por institución y curvas de potencia del diseño **785**

ÍNDICE DE TABLAS

Tabla 1-1
Operacionalización de los constructos del modelo **60**

Tabla 1-2
Cálculo a priori del tamaño de la muestra por análisis **61**

Tabla 1-3
Propiedades psicométricas del modelo de medición (análisis factorial confirmatorio) **63**

Tabla 1-4
Índices de ajuste del modelo hipotetizado y del modelo alternativo unifactorial **64**

Tabla 1-5
Validez discriminante: criterio de Fornell-Larcker y razón HTMT **64**

Tabla 1-6
Estadísticos descriptivos y correlaciones entre constructos **65**

Tabla 1-7
Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis **68**

Tabla 1-8
Descomposición de los efectos de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora **69**

Tabla 1-10
Estabilidad del efecto directo de la diversificación epistémica formativa frente a perturbaciones de la muestra (puntuaciones compuestas) **70**

Tabla 1-9
Regresión jerárquica de la práctica pedagógica descolonizadora sobre puntuaciones compuestas (análisis de robustez) **73**

Tabla 2-1
Operacionalización de las variables (diseño de medición) **97**

Tabla 2-2
Cálculo a priori del tamaño muestral por análisis **98**

Tabla 2-3
Modelo de medición: propiedades psicométricas por ítem **100**

Tabla 2-4
Validez discriminante (criterio de Fornell-Larcker y HTMT) **100**

Tabla 2-5
Estadísticos descriptivos y correlaciones **101**

Tabla 2-6
Ajuste global y comparación de modelos (CB-SEM) **101**

Tabla 2-7
Rutas estructurales del modelo CB-SEM **105**

Tabla 2-8
Descomposición de efectos (mediación, bootstrap 5000 réplicas) **105**

Tabla 2-9
Estabilidad del efecto total (jackknife, submuestreo y bootstrap) **108**

Tabla 3-1
Operacionalización de las variables **129**

Tabla 3-2
Modelo de medición: propiedades psicométricas por ítem **131**

Tabla 3-3
Validez discriminante: HTMT y criterio de Fornell-Larcker **132**

Tabla 3-4
Estadísticos descriptivos y correlaciones **132**

Tabla 3-5
Ajuste global del modelo e invarianza de medición entre zonas **133**

Tabla 3-6

Rutas estructurales del modelo CB-SEM **133**

Tabla 3-7
Descomposición de efectos de la mediación **137**

Tabla 3-8
Comparación de los constructos entre zonas **137**

Tabla 3-9
Análisis multigrupo del efecto de la gestión sobre la calidad del aprendizaje **141**

Tabla 3-10
Estabilidad del efecto total de la gestión bajo perturbaciones de la muestra **141**

Tabla 4-1
Ítems del instrumento y propiedades del modelo de medida **163**

Tabla 4-2
Cálculo a priori del tamaño de la muestra **164**

Tabla 4-3
Estadísticos descriptivos y correlaciones entre constructos **166**

Tabla 4-4
Validez discriminante: razón heterorrasgo-monorrasgo y criterio de Fornell-Larcker **166**

Tabla 4-5
Índices de ajuste del modelo hipotetizado y del modelo alternativo **170**

Tabla 4-6
Coeficientes estructurales del modelo **175**

Tabla 4-7
Descomposición del efecto de la gobernanza escolar sobre el rendimiento académico **175**

Tabla 4-8
Varianza explicada y tamaño del efecto por ruta **180**

Tabla 4-9
Sensibilidad de los efectos al estimador **181**

Tabla 4-10
Estabilidad del efecto total ante el remuestreo de la muestra **181**

Tabla 5-1
Operacionalización de los constructos del modelo **201**

Tabla 5-2
Estadísticos descriptivos y correlaciones entre constructos **203**

Tabla 5-3
Modelo de medición: cargas, fiabilidad y validez convergente **204**

Tabla 5-4
Validez discriminante: criterio de Fornell-Larcker y razón HTMT **205**

Tabla 5-5
Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor **205**

Tabla 5-6
Rutas estructurales directas del modelo **209**

Tabla 5-7
Descomposición del efecto del conocimiento del razonamiento algebraico temprano sobre la integración **210**

Tabla 5-8
Regresión jerárquica de la integración del pensamiento algebraico en la enseñanza de las ciencias **213**

Tabla 5-9
Pruebas de estrés del coeficiente del conocimiento del razonamiento algebraico temprano (modelo M2) **214**

Tabla 5-10
Trazabilidad de las hipótesis: contraste, procedimiento y decisión **218**

Tabla 6-1
Cálculo a priori del tamaño muestral para los análisis del estudio **239**

Tabla 6-2
Perfil de los participantes por grupo disciplinar **241**

Tabla 6-3
Redacción de los indicadores, estadísticos descriptivos, cargas externas y colinealidad **242**

Tabla 6-4
Fiabilidad de consistencia interna, validez convergente y colinealidad completa **244**

Tabla 6-5
Validez discriminante: razones heterorrasgo-monorrasgo y criterio de Fornell–Larcker **245**

Tabla 6-6
Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis **250**

Tabla 6-7
Poder explicativo y capacidad predictiva fuera de muestra (PLSpredict) **256**

Tabla 6-8
Descomposición de los efectos de los antecedentes sobre la intención de uso indebido en directos, indirectos y totales **260**

Tabla 6-9
Comprobaciones de robustez: efectos no lineales y endogeneidad **261**

Tabla 6-10
Invarianza de medida de los compuestos (MICOM) y análisis multigrupo por grupo disciplinar **262**

Tabla 7-1
Operacionalización de los constructos del modelo **288**

Tabla 7-2
Modelo de medición: propiedades psicométricas por ítem **289**

Tabla 7-3
Validez discriminante **290**

Tabla 7-4
Ajuste global del modelo e invarianza de medición por lengua materna **291**

Tabla 7-5
Estadísticos descriptivos y correlaciones **295**

Tabla 7-6
Rutas estructurales del modelo CB-SEM **296**

Tabla 7-7
Descomposición de los efectos (mediación) **296**

Tabla 7-8
Robustez: regresión jerárquica por mínimos cuadrados ordinarios con covariables **300**

Tabla 7-9
Robustez: especificaciones alternativas) **301**

Tabla 7-10
Robustez: estabilidad del efecto de la exposición ante perturbaciones de la muestra **301**

Tabla 8-1
Operacionalización de los constructos **324**

Tabla 8-2
Características de los participantes por sexo **326**

Tabla 8-3
Comprensión conceptual al inicio y al cierre de la unidad de mecánica (porcentaje de aciertos) **327**

Tabla 8-4
Estadísticos descriptivos, cargas factoriales, fiabilidad y validez convergente **330**

Tabla 8-5
Validez discriminante: correlaciones entre factores, raíz de la AVE y razón HTMT **331**

Tabla 8-6
Índices de ajuste del modelo de medida, del modelo hipotetizado y de los modelos alternativos **332**

Tabla 8-7
Coeficientes del modelo estructural hipotetizado **336**

Tabla 8-8
Descomposición del efecto de las estrategias de compromiso interactivo sobre la comprensión conceptual final **337**

Tabla 8-9
Invarianza de medida entre mujeres y varones **341**

Tabla 8-10
Análisis multigrupo de las trayectorias estructurales por sexo **341**

Tabla 9-1
Estructura de la intervención con herramientas digitales y de la enseñanza habitual del grupo de comparación **365**

Tabla 9-2
Características de los participantes y equivalencia inicial entre condiciones (N = 543) **367**

Tabla 9-3
Cargas factoriales estandarizadas del análisis factorial confirmatorio de la prueba de razonamiento lógico-matemático en el pretest y el postest (N = 543) **368**

Tabla 9-4
Fiabilidad por dimensión y ajuste del modelo de medida de tres factores frente al modelo unidimensional **369**

Tabla 9-5
Estadísticos descriptivos del razonamiento lógico-matemático por condición y medición **373**

Tabla 9-6
Verificación de los supuestos del análisis de covarianza por variable dependiente **374**

Tabla 9-7
Análisis multivariado y univariados de covarianza del efecto de la intervención sobre el postest, controlando el pretest **375**

Tabla 9-8
Medias ajustadas por el pretest, diferencia ajustada, tamaño del efecto y ganancia normalizada por condición **379**

Tabla 9-9
Análisis de sensibilidad del efecto de la intervención con estimadores e inferencias alternativos **380**

Tabla 9-10
Distribución de los niveles de logro del puntaje total en el pretest y el postest por condición **381**

Tabla 10-1
Operacionalización de las variables del estudio **407**

Tabla 10-2
Modelo de medición: propiedades psicométricas por ítem **409**

Tabla 10-3
Validez discriminante según la razón HTMT y el criterio de Fornell y Larcker **410**

Tabla 10-4
Invarianza de medida entre mujeres y varones **410**

Tabla 10-5
Estadísticos descriptivos y correlaciones **414**

Tabla 10-6
Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor **415**

Tabla 10-7
Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis **416**

Tabla 10-8
Descomposición de efectos del aprendizaje autorregulado sobre la ansiedad matemática **416**

Tabla 10-9
Regresión jerárquica de la ansiedad matemática sobre las covariables y los predictores (MCO con errores HC3) **420**

Tabla 10-10
Estabilidad del efecto del aprendizaje autorregulado sobre la ansiedad matemática ante perturbaciones de la muestra **424**

Tabla 11-1
Características iniciales de los estudiantes analizados según condición **446**

Tabla 11-2
Fiabilidad y concordancia entre evaluadores de la prueba de resolución de problemas **447**

Tabla 11-3
Puntuaciones de resolución de problemas por condición y ocasión **448**

Tabla 11-4
Diagnósticos de los supuestos del análisis de covarianza **449**

Tabla 11-5
Análisis de covarianza de dos niveles de la puntuación total del postest **450**

Tabla 11-6
Análisis de covarianza de dos niveles por fase de resolución de problemas **454**

Tabla 11-7
Robustez del efecto de la condición sobre la puntuación total del postest **458**

Tabla 11-8
Proporción de estudiantes que alcanzan el dominio (total ≥ 39 de 64) según condición **459**

Tabla 11-9
Parámetros de diseño planificados y observados **463**

Tabla 12-1
Cálculo a priori del tamaño de la muestra **487**

Tabla 12-2
Ítems y propiedades del modelo de medida **488**

Tabla 1-3
Propiedades psicométricas del modelo de medición (análisis factorial confirmatorio) **489**

Tabla 12-3
Estadísticos descriptivos y correlaciones entre constructos **490**

Tabla 12-4
Validez discriminante: razón heterorrasgo-monorrasgo y criterio de Fornell-Larcker **494**

Tabla 12-5
Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor **495**

Tabla 12-6
Coeficientes estructurales del modelo **495**

Tabla 12-7
Descomposición del efecto de la experiencia de excursión sobre el rendimiento académico **499**

Tabla 12-8
Varianza explicada y tamaño del efecto por ruta **499**

Tabla 12-9
Sensibilidad de los efectos al estimador **503**

Tabla 12-10
Estabilidad del efecto total de la experiencia de excursión sobre el rendimiento académico **504**

Tabla 13-1
Cálculo a priori del tamaño muestral (método de la raíz cuadrada inversa) **530**

Tabla 13-2
Operacionalización de los constructos y de la variable de agrupación **531**

Tabla 13-3
Modelo de medición: cargas externas, fiabilidad y validez convergente **533**

Tabla 13-4
Estadísticos descriptivos, correlaciones entre constructos y validez discriminante **533**

Tabla 13-5
Coeficientes de ruta del modelo estructural PLS-SEM **538**

Tabla 10-8
Descomposición de efectos: directo, indirecto y total **538**

Tabla 13-7
Ajuste aproximado, capacidad explicativa y capacidad predictiva fuera de muestra **542**

Tabla 13-8
Sesgo de método común: factores de inflación de la varianza de colinealidad completa **542**

Tabla 13-9
Invarianza de medición de los compuestos entre tipos de institución (MICOM) **547**

Tabla 13-10
Análisis multigrupo PLS por permutación: universidades frente a escuelas de educación superior pedagógica **548**

Tabla 14-1
Cálculo a priori del tamaño de la muestra para los tres contrastes del estudio (δ = .05; potencia objetivo = .80) **573**

Tabla 14-2
Operacionalización de los constructos, descriptivos por ítem y fiabilidad entre evaluadores (N = 2 422) **575**

Tabla 14-3
Características de los participantes y equivalencia inicial entre condiciones (N = 2 422) **578**

Tabla 10-2
Modelo de medida: cargas factoriales estandarizadas, fiabilidad compuesta y varianza media extraída por oleada (N = 2 422) **582**

Tabla 14-5
Invarianza de medida longitudinal (pretest-postest) y entre condiciones (tertulias dialógicas frente a comparación) **583**

Tabla 14-6
Validez discriminante: raíz cuadrada de la varianza media extraída (diagonal), correlaciones latentes (bajo la diagonal) y razón heterorrasgo-monorrasgo (sobre la diagonal) **583**

Tabla 14-7
Descriptivos por condición y oleada, y diferencia ajustada en el postest (N = 2 422) **584**

Tabla 14-8
Modelo de ecuaciones estructurales latente de dos oleadas: coeficientes estructurales (N = 2 422) **585**

Tabla 14-9
Descomposición del efecto de las tertulias dialógicas sobre la lectura crítica: efectos total, directo e indirecto a través de la empatía literaria intercultural **587**

Tabla 14-10
Análisis de robustez y de heterogeneidad del efecto **588**

Tabla 15-1
Variables, instrumentos y puntuación **612**

Tabla 15-2
Características basales y equivalencia de la muestra analítica (N = 990; 62 aulas) **614**

Tabla 15-3
Fiabilidad y estadísticos de los ítems de la prueba de comprensión (N = 990) **615**

Tabla 15-4
Puntajes de pretest, postest y cambio por brazo (N = 990) **616**

Tabla 15-5
Verificación de supuestos y agrupamiento de los resultados del postest (N = 990; 62 aulas) **617**

Tabla 15-6
Análisis de covarianza multinivel de la comprensión inferencial del postest (N = 990; 62 aulas) **618**

Tabla 15-7
Efectos grupo × tiempo: modelos multinivel de cambio pretest–postest (N = 990; 62 aulas) **619**

Tabla 15-8
Efectos del programa por tipo de inferencia, prueba multivariada y contraste de especificidad **620**

Tabla 15-9
Tamaños del efecto estandarizados con intervalos de confianza bootstrap por conglomerados **621**

Tabla 15-10
Robustez del efecto del programa sobre la comprensión inferencial del postest **623**

Tabla 16-1
Operacionalización de los constructos del estudio **645**

Tabla 16-2
Cálculo a priori del tamaño de la muestra **648**

Tabla 16-3
Composición de la muestra y retención por región **650**

Tabla 16-4
Estadísticos descriptivos, correlación intraclase y correlaciones **651**

Tabla 16-5
Modelo de medida longitudinal y validez discriminante **652**

Tabla 16-6
Invarianza longitudinal de la medida de comprensión lectora **653**

Tabla 16-7
Comparación de modelos de crecimiento latente por lengua **655**

Tabla 16-8
Modelo de crecimiento latente en paralelo condicional **656**

Tabla 16-9
Dinámica intrapersona entre la comprensión en lengua originaria y en castellano **659**

Tabla 16-10
Análisis de robustez de los efectos focales **661**

Tabla 17-1
Características de los participantes y equivalencia inicial por condición (N = 432) **684**

Tabla 17-2
Operacionalización de las medidas y fiabilidad **685**

Tabla 17-3
Estadísticos descriptivos de la precisión sintáctica por condición y medición **686**

Tabla 17-4
Invarianza longitudinal del modelo de medida de la precisión sintáctica (n = 389) **686**

Tabla 17-5
Modelos lineales mixtos de la precisión sintáctica: pruebas de Wald de los efectos fijos y componentes de varianza **687**

Tabla 17-6
Contrastes planificados: diferencia entre condiciones en el cambio desde el pretest **688**

Tabla 17-7
Cambio en la tasa de errores por categoría: razones de tasas de incidencia relativas al control **690**

Tabla 17-8
Modelo de ecuaciones estructurales del mecanismo: condición → compromiso con la retroalimentación → precisión sintáctica posterior (n = 419) **692**

Tabla 17-9
Moderación del efecto en el postest (CSE/C) por la competencia inicial en inglés **694**

Tabla 17-10
Robustez del efecto primario (RIA – C, postest, CSE/C) ante estimadores e inferencias alternativos **695**

Tabla 18-1
Cálculo a priori del tamaño muestral por criterio **719**

Tabla 18-2
Características de la muestra por lengua originaria (N = 480 estudiantes; J = 32 instituciones) **722**

Tabla 18-3
Indicadores, estadísticos descriptivos y cargas factoriales estandarizadas (N = 480) **723**

Tabla 18-4
Fiabilidad, validez convergente y dependencia por institución de los constructos **724**

Tabla 18-5
Validez discriminante: raíz cuadrada de la VME frente a las correlaciones entre factores y razón heterorrasgo-monorrasgo (HTMT) **724**

Tabla 18-6
Ajuste de los modelos de medición y estructurales rivales (N = 480) **725**

Tabla 18-7
Trayectorias estructurales del modelo hipotetizado (M5) **726**

Tabla 18-8
Descomposición de los efectos sobre el desempeño en inglés (coeficientes estandarizados) **728**

Tabla 18-9
Invarianza de la medición y de las trayectorias estructurales entre estudiantes quechuahablantes (n = 240) y aimarahablantes (n = 240) **729**

Tabla 18-10
Comparación de las trayectorias estructurales y de los efectos indirectos entre grupos de lengua originaria (modelo con invarianza escalar) **730**

INTRODUCCIÓN

En la universidad, la educación se torna competitiva entre los estudiantes, pues estos se forman para desenvolverse adecuadamente en los contextos laborales. Asimismo, en la educación universitaria se implican diferentes factores que logran influir en el aprendizaje óptimo de los estudiantes, pues el rendimiento académico suele ser un parámetro sensible a los cambios sociales y personales. Así, el aprendizaje de los estudiantes universitarios debe ser una prioridad por parte de las autoridades y de la institución universitarias para asegurar una calidad de enseñanza-aprendizaje que mejore la asimilación de contenidos académicos.

En la educación, es posible hallar diferentes teorías de adquisición de contenidos. Según Moreira (2017), la teoría del aprendizaje significativo resulta ser necesaria y actual para la educación universitaria, pues en las prácticas contemporáneas de la pedagogía existe un predominio de las pruebas mecánicas, lo que no permite una adquisición de contenido significativo para cada alumno. Asimismo, esta teoría puede ser complementada por otras teorías y enfoques educativos a fin de mejorar la enseñanza centrada en el placer por aprender, el significado de lo aprendido y en la compresión.

Por su parte, Michelena (2016) señala que la educación actual precisa de una postura desde la teoría de aprendizaje constructivista, pues de esta forma es posible observar el desarrollo genuino de los estudiantes a través de una construcción propia devenida de sus relaciones sociales con la sociedad y con su ambiente, de tal forma que se asimila la experiencia

desde una perspectiva propia. Así, la educación resulta ser una forma más de interacción donde se inmiscuye aspectos sociales, afectivos y cognitivos.

El aprendizaje resulta ser un aspecto importante de la educación porque mediante ese se observa si los estudiantes realmente están adquiriendo sus conocimientos académicos. No obstante, se debe procurar una buena enseñanza a nivel universitario para que el aprendizaje también sea óptimo. Los centros de educación superior deben contar con las disposiciones y los recursos necesarios que aseguren una buena enseñanza por parte de los docentes, pues estos son responsables directo de que todo aprendizaje sea dinámica y didáctica.

Según De Pablos et al. (2019), las universidades son sistemas educativos importantes para desarrollar todo conocimiento científico que permite la formación de profesionales con niveles óptimos. Es por ello que se procura por la vanguardia en cuanto a la actualización de recursos educativos como la transformación y la innovación digitales, la cual se integra en la enseñanza y sus modelos. Es decir, la enseñanza se torna mejor cuando se emplea diferentes herramientas con cuales los estudiantes sean capaces de aprender mejor a niveles óptimos.

Respecto a la enseñanza de las matemáticas en las universidades, las actitudes suelen variar por diferentes factores. Según señalan Flores y Auzmendi (2018), en la enseñanza de las matemáticas se incurren variables como el sexo y la etnia que influyen en las actitudes para el aprendizaje de todo concepto matemático, incluso en el lenguaje científico. A partir de esto, se puede inferir que es importante aplicar diferentes estrategias didácticas para optimizar el aprendizaje de las matemáticas.

Es así que la aplicación de manuales de reglas para mejorar algunos temas matemáticos, como los sistemas de unidades, se posibilita como una opción viable para el incremento de los niveles de dominio y uso tanto del tema matemático de los sistemas de unidades como el manejo del lenguaje científico en este ámbito científico.

CAPÍTULO I

DIVERSIFICACIÓN EPISTÉMICA DE LA FORMACIÓN DOCENTE Y PRÁCTICA PEDAGÓGICA DESCOLONIZADORA EN HUMANIDADES: EL PAPEL MEDIADOR DE LA REFLEXIVIDAD EPISTÉMICA CRÍTICA EN EL PERÚ

Wilder Antonio Vega Alcántara

Universidad Nacional del Centro del Perú

Resumen

La formación docente decide qué saberes se legitiman, pero se desconoce el mecanismo por el cual la diversificación epistémica de esa formación se traduce en práctica pedagógica descolonizadora. Este estudio examinó en qué medida dicha diversificación se asocia con la práctica descolonizadora de docentes de humanidades del Perú y si la reflexividad epistémica crítica transmite esa asociación. Con un diseño cuantitativo, transversal y explicativo, se estimó por máxima verosimilitud un modelo de ecuaciones estructurales basado en covarianzas con tres constructos latentes reflectivos (nueve ítems) en una muestra por conveniencia de 375 docentes, cuyo tamaño se fijó a priori mediante la potencia de la prueba de ajuste cercano. El modelo de medición mostró cargas de .563 a .885, fiabilidad adecuada y un ajuste satisfactorio (CFI = 1.000; RMSEA = .000; SRMR = .024) que superó al modelo unifactorial. La diversificación se asoció con la práctica de forma directa (β = .309) y con la reflexividad (β = .790); el efecto indirecto fue significativo (a × b = .238; IC 95 % bootstrap [.059, .440]), de modo que la mediación fue parcial y explicó el 43.5 % del efecto total (.547). Una regresión jerárquica y diversas pruebas de estabilidad replicaron la dirección de los resultados. Los hallazgos son compatibles con un mecanismo de conversión entre formación y práctica, aunque el diseño transversal, el autorreporte y la cercanía entre constructos acotan la inferencia. Se discuten implicaciones para la formación docente intercultural.

Palabras clave: pedagogía descolonial, formación docente, reflexividad epistémica, mediación, ecuaciones estructurales, humanidades, Perú.

Abstract

Teacher education shapes which knowledges are recognized as legitimate, yet the mechanism through which its epistemic diversification becomes decolonizing pedagogical practice remains poorly specified. This study examined the extent to which epistemic diversification in teacher education is associated with the

decolonizing practice of humanities teachers in Peru, and whether critical epistemic reflexivity transmits that association. Using a quantitative, cross-sectional, explanatory design, a covariance-based structural equation model with three reflective latent constructs (nine items) was estimated by maximum likelihood in a convenience sample of 375 teachers, whose size was set a priori through the power of the close-fit test. The measurement model showed loadings between .563 and .885, adequate reliability, and satisfactory fit (CFI = 1.000; RMSEA = .000; SRMR = .024) that clearly outperformed the one-factor alternative. Diversification was associated with practice directly (β = .309) and with reflexivity (β = .790); the indirect effect was significant (a × b = .238; 95% bootstrap CI [.059, .440]), indicating partial mediation that accounted for 43.5% of the total effect (.547). A hierarchical regression and several stability checks replicated the direction of the results. The findings are consistent with a conversion mechanism between training and practice, although the cross-sectional design, self-report measures, and the proximity between constructs limit inference. Implications for intercultural teacher education are discussed.

Keywords*: *decolonial pedagogy, teacher education, epistemic reflexivity, mediation, structural equation modeling, humanities, Peru.

1.1 INTRODUCCIÓN

La formación docente constituye uno de los espacios en los que se decide qué conocimientos se reconocen como legítimos y cuáles quedan relegados al lugar de lo anecdótico o lo folclórico. En América Latina, esa decisión arrastra la herencia de la colonialidad del poder y del saber, entendida como la clasificación jerárquica de poblaciones y conocimientos que hizo del eurocentrismo la racionalidad dominante de la modernidad (Quijano, 2000; Gandarilla et al., 2021). La crítica epistemológica del Sur ha mostrado que esa jerarquía produce una injusticia cognitiva: la incapacidad de reconocer las distintas formas de conocer mediante las cuales las personas dan sentido a su existencia (Santos, 2014). En el campo educativo, la injusticia epistémica asociada al currículo ha dejado de ser una

preocupación exclusivamente filosófica para convertirse en una urgencia pedagógica, porque los currículos vigentes generan consecuencias epistémicas negativas para estudiantes cuyas tradiciones de conocimiento no aparecen en ellos (Akbar, 2024; Stopford, 2025; Zembylas, 2025).

AI Mode conversation: Redacta el siguiente texto tomando siempre que sea posible como criterio para hacerlo reducir verbos, es decir en vez de decir "no exige" decir "exime". Este es el texto: "El Perú ofrece un sitio de prueba especialmente exigente para esta discusión. Su sistema educativo reconoce formalmente la diversidad lingüística y cultural del país y ha institucionalizado la Educación Intercultural Bilingüe (EIB) en la educación básica (Flores et al., 2023). Sin embargo, la evidencia reciente muestra que la diversidad ocupa con frecuencia un lugar marginal en la vida escolar, como ocurre con el uso de las lenguas originarias, restringido a pequeños espacios del aula mientras el castellano domina el resto de la escuela (Sullón, 2024). En la formación inicial docente, el nuevo currículo de formación para la EIB ha abierto oportunidades de innovación, pero sus implementadores advierten limitaciones para abordar la diversidad cultural, una colaboración interinstitucional débil y una alta rotación de formadores (Moreano et al., 2026). La tensión no es exclusiva del Perú: el análisis del currículo de formación docente en otros sistemas muestra que los programas formales pueden perpetuar posiciones de desconocimiento frente a los saberes indígenas, aun cuando el discurso institucional declare apertura (Mattila et al., 2025), y que los docentes indígenas perciben la descolonización del currículo como una tarea pendiente más que como una realidad consolidada (Bustos et al., 2026)."

Redacta el siguiente texto tomando siempre que sea posible como criterio para hacerlo reducir verbos, es decir en vez de decir "no exige" decir "exime". Este es el texto: "El Perú ofrece un sitio de prueba especialmente exigente para esta discusión. Su sistema educativo reconoce formalmente la diversidad lingüística y cultural del país y ha institucionalizado la Educación Intercultural Bilingüe (EIB) en la educación básica (Flores et al., 2023). Sin embargo, la evidencia reciente muestra que la diversidad ocupa con frecuencia un lugar marginal en la vida escolar, como

ocurre con el uso de las lenguas originarias, restringido a pequeños espacios del aula mientras el castellano domina el resto de la escuela (Sullón, 2024). En la formación inicial docente, el nuevo currículo de formación para la EIB ha abierto oportunidades de innovación, pero sus implementadores advierten limitaciones para abordar la diversidad cultural, una colaboración interinstitucional débil y una alta rotación de formadores (Moreano et al., 2026). La tensión no es exclusiva del Perú: el análisis del currículo de formación docente en otros sistemas muestra que los programas formales pueden perpetuar posiciones de desconocimiento frente a los saberes indígenas, aun cuando el discurso institucional declare apertura (Mattila et al., 2025), y que los docentes indígenas perciben la descolonización del currículo como una tarea pendiente más que como una realidad consolidada (Bustos et al., 2026).”

El Perú ofrece un escenario de prueba sumamente exigente para esta discusión. Su sistema educativo reconoce formalmente la diversidad lingüística y cultural del país y ha institucionalizado la Educación Intercultural Bilingüe (EIB) en la educación básica (Flores et al., 2023). Sin embargo, la evidencia reciente muestra que la diversidad ocupa con frecuencia un lugar marginal en la vida escolar; por ejemplo, el uso de lenguas originarias queda restringido a pequeños espacios del aula, mientras el castellano domina el resto de la escuela (Sullón, 2024). En la formación inicial docente, el nuevo currículo para la EIB ha abierto oportunidades de innovación, pero sus implementadores advierten limitaciones para abordar la diversidad cultural, una débil colaboración interinstitucional y una alta rotación de formadores (Moreano et al., 2026). Esta tensión trasciende al Perú: el análisis curricular de la formación docente en otros sistemas muestra que los programas formales perpetúan el desconocimiento de los saberes indígenas, pese a la apertura del discurso institucional (Mattila et al., 2025), y que los docentes indígenas perciben la descolonización curricular como una tarea pendiente antes que como una realidad consolidada (Bustos et al., 2026).

La literatura sobre descolonización pedagógica ha sido predominantemente conceptual y cualitativa. Ha documentado la genealogía colonial de los currículos, ha propuesto marcos para descentrar la enseñanza de las humanidades (R'boul et al., 2026) y la internacionalización de la educación superior desde epistemologías del Sur (Thakur y Mukherjee, 2026), y ha explorado la reflexividad como condición para una práctica descolonizadora (Allen y Girei, 2024). Paralelamente, la investigación sobre creencias docentes y enseñanza culturalmente pertinente ha mostrado que las disposiciones de los docentes no se traducen de manera automática en su práctica, y que la autoeficacia y la formación profesional orientada a la equidad intervienen de forma desigual en esa traducción. Estos hallazgos sugieren que entre lo que la formación ofrece y lo que el docente hace en el aula media un proceso que la literatura descolonial ha descrito, pero que pocas veces se ha modelado.

El vacío que aborda este estudio es la ausencia de evidencia cuantitativa sobre el mecanismo por el cual la diversificación epistémica de la formación docente —el grado en que la formación incorpora saberes plurales y presenta el canon académico como una perspectiva entre otras— se traduce en prácticas pedagógicas descolonizadoras en las humanidades. Se propone que ese mecanismo es la reflexividad epistémica crítica, es decir, la capacidad del docente de examinar el lugar epistémico desde el que enseña. La pregunta de investigación es: ¿en qué medida la diversificación epistémica de la formación docente se asocia con la práctica pedagógica descolonizadora de docentes de humanidades en el Perú, y en qué medida esa asociación se transmite a través de la reflexividad epistémica crítica? En consecuencia, se contrastan tres hipótesis, que se fundamentan en el marco teórico: H1, un efecto directo positivo de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora; H2, un efecto positivo de la diversificación epistémica formativa sobre la reflexividad epistémica crítica; y H3, un efecto indirecto positivo de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora a través de la reflexividad epistémica crítica.

El estudio aporta a la teoría de la pedagogía descolonial al especificar y contrastar empíricamente un mecanismo de conversión entre formación y práctica, y aporta metodológicamente al demostrar que constructos de raíz descolonial pueden medirse con validez y modelarse mediante ecuaciones estructurales basadas en covarianzas, sin reducirlos a categorías residuales. El caso peruano funciona como sitio de prueba de un argumento de alcance general para los sistemas educativos que declaran la interculturalidad en sus currículos.

1.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

El marco principal del estudio es la pedagogía crítica descolonial, que entiende el canon formativo no como un repertorio neutro de contenidos sino como un dispositivo que jerarquiza formas de conocer (Quijano, 2000; Santos, 2014). Desde esta perspectiva, descolonizar la pedagogía no consiste en añadir contenidos locales a un currículo intacto, sino en sostener los sistemas de conocimiento en tensión y disputar la autoridad epistémica del canon (R'boul et al., 2026). La diversificación epistémica de la formación docente designa, por tanto, la medida en que la formación expone al futuro docente a una pluralidad de tradiciones de conocimiento y desnaturaliza la universalidad del canon académico.

El marco complementario es la teoría de la reflexividad docente. La reflexividad media entre las condiciones contextuales y la agencia profesional de los formadores de docentes (Hinostroza, 2023), y la meta-reflexividad, entendida como la capacidad de hacer explícitos los fundamentos de distintas teorías y enfoques, se ha propuesto como un componente esencial del profesionalismo docente (Cramer et al., 2023). En la vertiente descolonial, la reflexividad exige que el docente confronte su propia posición dentro de las jerarquías epistémicas para no reproducirlas en su práctica (Allen y Girei, 2024). La evidencia sobre pedagogías relacionales muestra, además, que cuando la formación permite a los futuros docentes construir una conexión auténtica con formas de conocer indígenas, aumentan su disposición y su confianza para incorporarlas en el currículo (Macdonald et al., 2024; Walker, 2023).

De la articulación de ambos marcos se derivan tres hipótesis. Si la diversificación epistémica de la formación modifica las condiciones de selección del saber, cabe esperar un efecto directo sobre la práctica: H1 postula que la diversificación epistémica formativa se asocia positivamente con la práctica pedagógica descolonizadora. Si la exposición a una pluralidad de tradiciones desnaturaliza el canon, cabe esperar también que active el examen del propio lugar epistémico: H2 postula que la diversificación epistémica formativa se asocia positivamente con la reflexividad epistémica crítica. Finalmente, si la reflexividad es el mecanismo proximal que convierte la diversificación en acción pedagógica, la asociación entre formación y práctica debe transmitirse, al menos en parte, a través de ella: H3 postula un efecto indirecto positivo de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora a través de la reflexividad epistémica crítica. El modelo resultante es una mediación simple con tres constructos latentes reflectivos, en la que la ruta directa y la ruta indirecta se estiman de forma simultánea.

1.3 MATERIALES Y MÉTODOS

1.3.1 Diseño

El estudio adoptó un enfoque cuantitativo, con diseño no experimental, transversal y explicativo, referido al año 2026, y tomó al docente individual como nivel de análisis. El modelo hipotetizado se contrastó mediante un modelo de ecuaciones estructurales basado en covarianzas (CB-SEM) con tres variables latentes reflectivas: diversificación epistémica formativa (exógena), reflexividad epistémica crítica (mediadora) y práctica pedagógica descolonizadora (criterio).

1.3.2 Instrumentos

Los tres constructos se midieron con escalas de autorreporte de tres ítems cada una, con respuesta tipo Likert de cinco puntos. Las escalas se construyeron a partir de las definiciones teóricas del apartado anterior y su estructura se contrastó mediante análisis factorial confirmatorio. La

definición de la reflexividad epistémica crítica tomó como referente las propuestas para capturar la reflexividad epistémica de los formadores de docentes que enseñan sobre, para y en la diversidad (L'Estrange et al., 2023). La Tabla 1-1 presenta la operacionalización de los constructos.

Tabla 1-1

Operacionalización de los constructos del modelo

Constructo	Rol en el modelo	Definición operacional	Ítems	Escala
Diversificación epistémica formativa	Exógena	Grado en que la formación docente recibida incorporó pensadores y saberes andinos, amazónicos y latinoamericanos, trató los saberes locales como conocimiento válido y presentó la teoría pedagógica clásica como una perspectiva entre otras	3	Likert de 5 puntos
Reflexividad epistémica crítica	Mediadora	Frecuencia con que el docente examina cómo su formación reproduce jerarquías entre formas de conocer, identifica sesgos eurocéntricos en su práctica y cuestiona el origen y la autoridad de los contenidos que enseña	3	
Masa	Criterio	Frecuencia con que el docente incorpora saberes comunales de sus estudiantes, diseña actividades que visibilizan la diversidad epistémica del Perú y evalúa reconociendo formas de conocer no occidentales	3	

Nota: Escala de respuesta de 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo). Los ítems completos y sus propiedades psicométricas se presentan en la Tabla 1-3.

1.3.3 Participantes y tamaño de la muestra

La población estuvo constituida por docentes del área de humanidades de la educación básica regular del Perú. Se empleó un muestreo no probabilístico por conveniencia, adecuado para un estudio explicativo orientado a contrastar relaciones entre constructos y no a estimar parámetros poblacionales; los resultados describen a la muestra analizada y no se afirma su generalización estadística. El tamaño muestral se determinó a priori con el método de potencia para la prueba de ajuste cercano

de MacCallum et al. (1996): con 24 grados de libertad (45 momentos muestrales menos 21 parámetros libres), un error cuadrático medio de aproximación de .05 bajo la hipótesis nula frente a .08 bajo la alternativa, $\alpha = .05$ y potencia de .80, se requieren como mínimo 375 participantes, número que se fijó como tamaño de la muestra. La potencia se obtuvo de la distribución chi-cuadrado no central, con parámetro de no centralidad $\lambda = (N - 1) \cdot gl \cdot \varepsilon^2$. Ese tamaño cubre los otros dos análisis del estudio (Tabla 2): para el efecto indirecto, una simulación Monte Carlo con intervalos bootstrap percentil al 95 % y rutas a = b = .26, valor intermedio entre efecto pequeño y mediano según Fritz y MacKinnon (2007), estimada con fiabilidad de .80; y para la prueba F global de la regresión del criterio sobre dos predictores, con $f^2 = .15$ (Hair et al., 2022). El análisis más exigente es, por tanto, el del ajuste global.

Tabla 1-2

Cálculo a priori del tamaño de la muestra por análisis

Análisis	Método y fuente	Parámetros	N mínimo	Potencia con	
				N mínimo	N = 375
Ajuste global del CB-SEM (prueba de ajuste cercano)	Chi-cuadrado no central; MacCallum et al. (1996)	$gl = 24$; $\varepsilon_0 = .05$; $\varepsilon_\alpha = .08$; $\alpha = .05$	375	.801s	.801
Efecto indirecto a × b	Simulación Monte Carlo, IC bootstrap percentil 95 % (1000 remuestreos, 1000 réplicas); Fritz y MacKinnon (2007)	a = b = .26; c′ = .14; fiabilidad .80	235	.806	.966
Prueba F global de la regresión del criterio	F no central; Hair et al. (2022)	$k = 2$; $f^2 = .15$; $\alpha = .05$	68	.804	> .999

Nota: gl = 45 momentos muestrales − 21 parámetros libres del modelo de medición y estructural. La N del estudio (N = 375) corresponde al análisis más exigente (ajuste global) y cubre los demás.

1.3.4 Análisis de datos

El análisis siguió dos etapas:

1) En la primera se evaluó el modelo de medición mediante análisis factorial confirmatorio: cargas factoriales estandarizadas (λ), varianza explicada de cada ítem (R^2), fiabilidad (α de Cronbach y ω de McDonald) y varianza media extraída (AVE). La validez discriminante se valoró con el criterio de Fornell y Larcker (1981) y con la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015), con umbrales de .85 y .90. El ajuste global se evaluó con χ^2, χ^2/gl, CFI, TLI, RMSEA y SRMR, y se comparó con un modelo alternativo unifactorial mediante la diferencia de χ^2; para el SRMR se adoptó el criterio < .08 (Kline, 2023).

2) En la segunda etapa se estimó el modelo estructural por máxima verosimilitud con el paquete semopy (Meshcheryakov et al., 2021). El efecto indirecto se contrastó con intervalos de confianza bootstrap percentil al 95 % con 5000 réplicas (Hayes, 2022); se consideró significativo cuando el intervalo excluía el cero. Como análisis de robustez, sin sustituir al modelo principal, se estimó una regresión jerárquica por mínimos cuadrados ordinarios sobre puntuaciones compuestas con errores estándar robustos HC3, y se evaluó la estabilidad del efecto directo mediante jackknife, submuestreo, bootstrap, winsorización del criterio y un predictor placebo obtenido por permutación.

1.4 RESULTADOS

1.4.1 Modelo de medición

Todas las cargas factoriales fueron significativas y superiores a .50, con valores entre .563 y .885 (Tabla 1-3). La fiabilidad resultó adecuada en los tres constructos y la varianza media extraída superó .50 en todos ellos. El ítem de menor carga fue «Diseño actividades que visibilizan la diversidad epistémica del Perú», que se conservó para no dejar la escala con dos ítems.

Tabla 1-3

Propiedades psicométricas del modelo de medición (análisis factorial confirmatorio)

Constructo	Ítem	λ	R^2	α	ω	AVE
Diversificación epistémica formativa	Mi formación docente incluyó pensadores andinos, amazónicos y latinoamericanos junto a los autores del canon europeo	.857	.734	.868	.868	.687
	El plan de estudios de mi formación trató los saberes locales como conocimiento válido	.771	.594	.868	.868	.687
	Mis formadores presentaron la teoría pedagógica clásica como una perspectiva entre otras, no como universal	.855	.731	.868	.868	.687
Práctica pedagógica descolonizadora	Incorporo saberes comunales de mis estudiantes en mis sesiones de humanidades	.749	.561	.781	.783	.554
	Diseño actividades que visibilizan la diversidad epistémica del Perú	.563	.317	.781	.783	.554
	Evalúo aprendizajes reconociendo formas de conocer no occidentales	.885	.783	.781	.783	.554
Reflexividad epistémica crítica	Analizo cómo mi propia formación reproduce jerarquías entre formas de conocer	.759	.576	.790	.790	.556
	Examino mis prácticas para identificar sesgos eurocéntricos	.719	.517	.790	.790	.556
	Cuestiono el origen y la autoridad de los contenidos que enseño	.759	.576	.790	.790	.556

Nota: N = 375. λ = carga factorial estandarizada; R^2 = varianza explicada del ítem; α = alfa de Cronbach; ω = omega de McDonald; AVE = varianza media extraída. Criterios: λ ≥ .50; α y ω ≥ .70; AVE ≥ .50.

El modelo hipotetizado mostró un ajuste global satisfactorio y superó con claridad al modelo alternativo de un solo factor (Tabla 4); la diferencia entre ambos fue significativa, $\Delta\chi^2(3) = 276.537$, $p < .001$. La parte estructural del modelo está saturada (tres rutas entre tres constructos), por lo que el ajuste global coincide con el del modelo de medición.

Tabla 1-4

Índices de ajuste del modelo hipotetizado y del modelo alternativo unifactorial

Índice	Modelo	
	hipotetizado (tres factores)	alternativo (un factor)
χ^2	20.093	296.630
gl	24	27
p (χ^2)	.691	< .001
χ^2/gl	0.837	10.986
CFI	1.000	0.831
TLI	1.000	0.775
RMSEA	0.000	0.163
SRMR	0.024	0.118
$\Delta\chi^2(\Delta gl)$ frente al modelo hipotetizado	—	276.537 (3), p < .001

Nota: N = 375. λ = carga factorial estandarizada; R^2 = varianza explicada del ítem; α = alfa de Cronbach; ω = omega de McDonald; AVE = varianza media extraída. Criterios: λ ≥ .50; α y ω ≥ .70; AVE ≥ .50.

Todos los valores HTMT quedaron por debajo del umbral estricto de .85, con un máximo de .810 entre diversificación epistémica formativa y reflexividad epistémica crítica (Tabla 1-5). Por el criterio de Fornell y Larcker, la diversificación epistémica formativa y la práctica pedagógica descolonizadora cumplieron el requisito, mientras que la reflexividad epistémica crítica no lo hizo: su raíz cuadrada de la AVE (.746) fue inferior a su correlación máxima con otro constructo (.771).

Tabla 1-5

Validez discriminante: criterio de Fornell-Larcker y razón HTMT

Constructo	1	2	3
1. Diversificación epistémica formativa	**.829**	.810	.578
2. Reflexividad epistémica crítica	.771	**.746**	.585
3. Práctica pedagógica descolonizadora	.527	.518	**.744**

Nota: Diagonal (negrita): raíz cuadrada de la AVE. Bajo la diagonal: correlaciones entre constructos (Tabla 6). Sobre la diagonal: razón heterorrasgo-monorrasgo (HTMT); criterio < .85 (estricto) y < .90 (liberal; Henseler et al., 2015). El criterio de Fornell-Larcker exige que la raíz de la AVE supere las correlaciones del constructo; no se cumple para la reflexividad epistémica crítica (.746 frente a .771).

La Tabla 1-6 presenta los estadísticos descriptivos y las correlaciones. Las tres medias se situaron en el punto medio de la escala, con desviaciones estándar entre 0.663 y 0.667 y asimetría y curtosis cercanas a cero. Todas las correlaciones fueron positivas y significativas, con la mayor entre diversificación y reflexividad. La Figura 1-1 sintetiza el modelo de medición.

Tabla 1-6

Estadísticos descriptivos y correlaciones entre constructos

	Constructo	M	DE	Asimetría	Curtosis	1	2	3
1.	Diversificación epistémica formativa	3.00	0.667	.209	−.207	—	.527*** [.449, .596]	.771*** [.727, .809]
2.	Práctica pedagógica descolonizadora	3.00	0.663	−.047	.033	.527***	—	.518*** [.440, .588]
3.	Reflexividad epistémica crítica	3.00	0.667	−.078	−.124	.771***	.518***	—

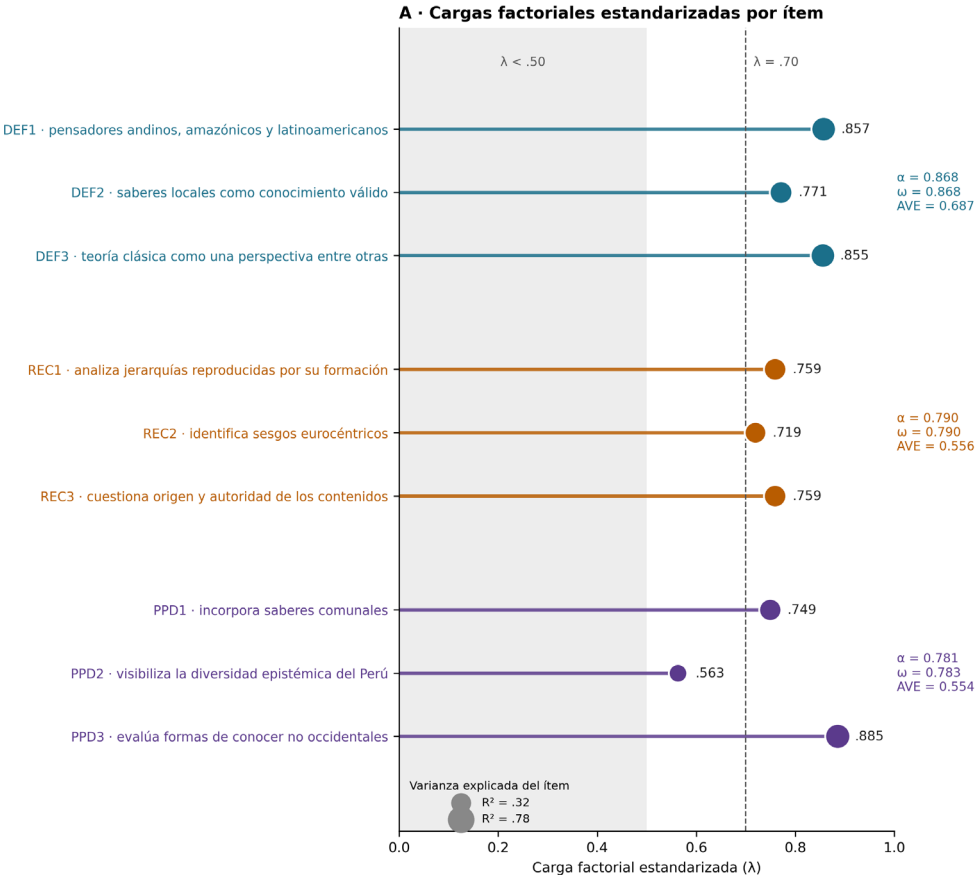
Nota: Diagonal (negrita): raíz cuadrada de la AVE. Bajo la diagonal: correlaciones entre constructos (Tabla 6). Sobre la diagonal: razón heterorrasgo-monorrasgo (HTMT); criterio < .85 (estricto) y < .90 (liberal; Henseler et al., 2015). El criterio de Fornell-Larcker exige que la raíz de la AVE supere las correlaciones del constructo; no se cumple para la reflexividad epistémica crítica (.746 frente a .771).

1.4.2 Modelo estructural y contraste de hipótesis

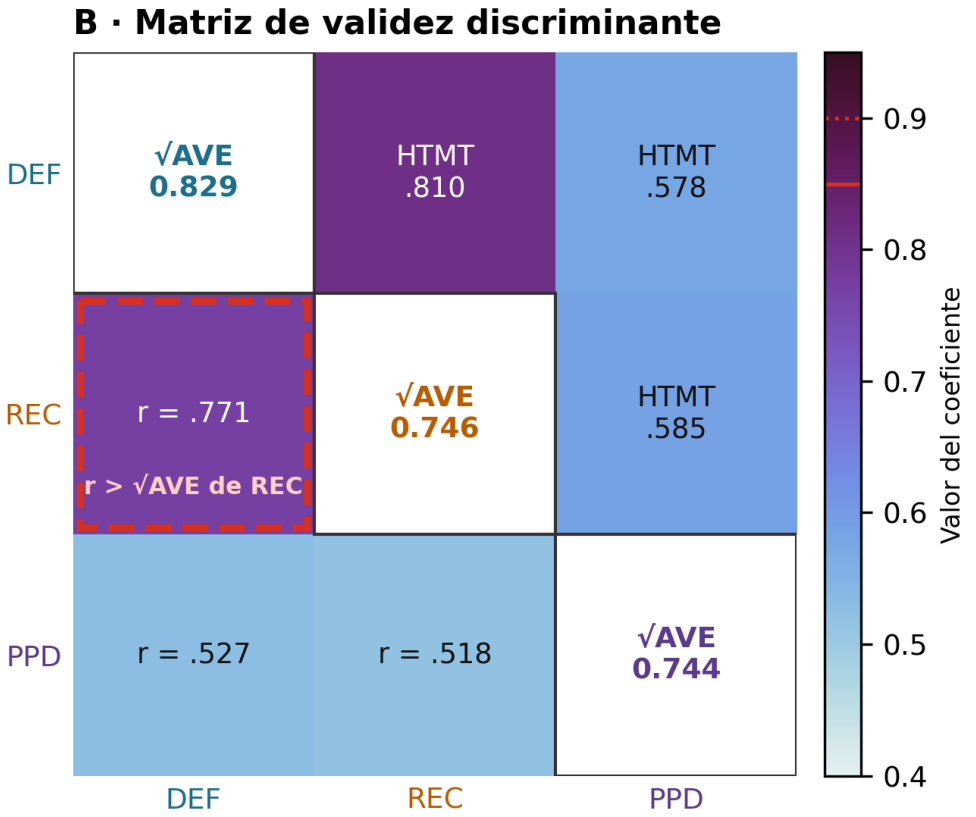
La Tabla 1-7 presenta las rutas estructurales estandarizadas. La diversificación epistémica formativa se asoció positivamente con la práctica pedagógica descolonizadora, por lo que H1 se soporta, y con la reflexividad epistémica crítica, por lo que H2 se soporta. La reflexividad epistémica crítica se asoció a su vez con la práctica pedagógica descolonizadora.

Figura 1-1

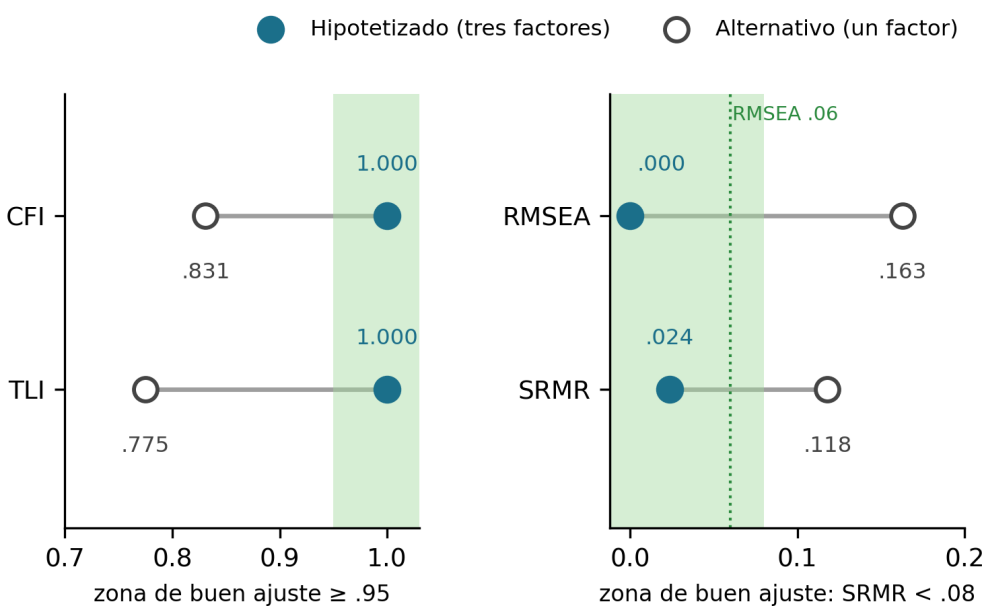
Modelo de medición: cargas factoriales, validez discriminante y ajuste global.



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel A: carga factorial estandarizada (λ) de cada uno de los nueve ítems, agrupados por constructo (DEF = diversificación epistémica formativa; REC = reflexividad epistémica crítica; PPD = práctica pedagógica descolonizadora); el tamaño del marcador es proporcional a la varianza explicada del ítem (R^2 entre .317 y .783), la zona gris marca λ < .50 y la línea discontinua λ = .70; a la derecha de cada bloque se indican α, ω y AVE del constructo (α entre .781 y .868; AVE entre .554 y .687). Las cargas van de .563 (PPD2) a .885 (PPD3). El texto íntegro de los ítems figura en la Tabla 1-3.



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel B: matriz mixta con la raíz cuadrada de la AVE en la diagonal (.829, .746 y .744), las correlaciones entre constructos bajo la diagonal (r = .771, .527 y .518) y la razón heterorrasgo-monorrasgo sobre ella (HTMT = .810, .578 y .585); las líneas rojas de la barra de color señalan los umbrales HTMT de .85 (continua) y .90 (punteada). La celda recuadrada indica el único incumplimiento del criterio de Fornell y Larcker: la correlación entre DEF y REC (.771) supera la raíz de la AVE de REC (.746), mientras que todos los HTMT quedan por debajo de .85 (Henseler et al., 2015).



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel C: índices de ajuste del modelo hipotetizado de tres factores (círculo relleno) frente al modelo alternativo de un factor (círculo vacío); las zonas verdes marcan CFI y TLI ≥ .95 y SRMR < .08, y la línea punteada RMSEA = .06. Modelo hipotetizado: $\chi^2(24) = 20.093$, p = .691, CFI = 1.000, TLI = 1.000, RMSEA = .000, SRMR = .024; modelo unifactorial: $\chi^2(27) = 296.630$, CFI = .831, TLI = .775, RMSEA = .163, SRMR = .118; $\Delta\chi^2(3) = 276.537$, p < .001.

Tabla 1-7

Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis

Ruta	Hipótesis	β	EE	p	IC 95 % bootstrap	Decisión
Diversificación epistémica formativa → Práctica pedagógica descolonizadora (c′)	H1	.309	.102	.008	[.072, .520]	Soportada
Diversificación epistémica formativa → Reflexividad epistémica crítica (a)	H2	.790	.056	< .001	[.724, .852]	Soportada
Reflexividad epistémica crítica → Práctica pedagógica descolonizadora (b)	—	.301	.120	.014	[.076, .544]	—

Ruta	Hipótesis	β	EE	p	IC 95 % bootstrap	Decisión
Efecto indirecto a × b	H3	.238	—	.010	[.059, .440]	Soportada

Nota: N = 375. Estimación por máxima verosimilitud (semopy). β = coeficiente estandarizado; EE = error estándar. Intervalos de confianza bootstrap percentil con 5000 réplicas; un efecto se considera significativo cuando su intervalo excluye el cero..

La descomposición de efectos (Tabla 1-8) mostró un efecto indirecto positivo cuyo intervalo excluyó el cero, por lo que H3 se soporta. El efecto directo se mantuvo significativo y el indirecto representó el 43.5 % del efecto total (.238 / .547), lo que configura una mediación parcial. La Figura 1-2 muestra el modelo completo y la descomposición.

Tabla 1-8

Descomposición de los efectos de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora

Efecto	Estimación	IC 95 %		Proporción del efecto total
		inferior	superior	
Ruta a (diversificación → reflexividad)	.790	.724	.852	—
Ruta b (reflexividad → práctica)	.301	.076	.544	—
Efecto directo c′	.309	.072	.520	56.5 %
Efecto indirecto a × b	.238	.059	.440	43.5 %
Efecto total c = c′ + a × b	.547	.450	.633	100.0 %

Nota: N = 375. Coeficientes estandarizados del modelo latente. Intervalos bootstrap percentil con 5000 réplicas (5000 convergentes). Mediación parcial: el efecto directo y el indirecto son significativos.

1.4.3 Soporte de hipótesis.

- H1 (efecto directo positivo): soportada.
- H2 (efecto positivo de la diversificación sobre la reflexividad): soportada.
- H3 (efecto indirecto positivo a través de la reflexividad): soportada, con un intervalo bootstrap que excluyó el cero.

1.4.4 Robustez

La regresión jerárquica sobre puntuaciones compuestas replicó la dirección y la significación de las rutas del modelo principal (Tabla 1-9). La incorporación de la reflexividad epistémica crítica en el Modelo 2 aumentó significativamente la varianza explicada ($\Delta R^2 = .031$, F de cambio(1, 372) = 16.51, $p < .001$) y redujo el coeficiente de la diversificación epistémica formativa. Los coeficientes de esta tabla son no estandarizados y proceden de otro estimador, por lo que no se comparan numéricamente con los β latentes de la Tabla 1-7.

Tabla 1-10

Estabilidad del efecto directo de la diversificación epistémica formativa frente a perturbaciones de la muestra (puntuaciones compuestas)

Prueba	Réplicas / n	b	DE o EE	IC 95 %	Signo estable (%)	Significativo (%)
Estimación de referencia (Modelo 1)	n = 375	.242	.020	[.203, .282]	—	—
Jackknife (exclusión de un caso)	375 réplicas	.242	.001	[.240, .245]	100 %	100 %
Submuestreo del 80 %	100 réplicas	.242	.010	[.223, .261]	100 %	100 %
Submuestreo del 60 %	100 réplicas	.240	.018	[.198, .273]	100 %	100 %
Bootstrap del efecto (remuestreo con reemplazo)	1000 réplicas	.242	.020	[.203, .282]	100 %	100 %
Criterio winsorizado al 5 %	n = 375	.212	.016	[.180, .244]	—	—
Predictor placebo (permutado)	n = 375	−.039	.024	[−.085, .008]	—	—

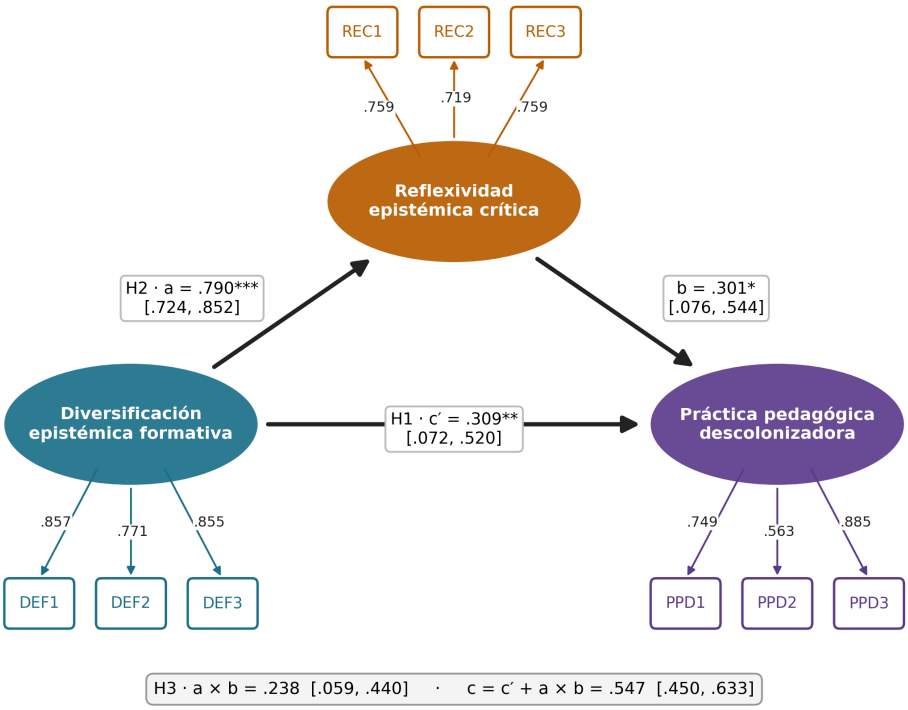
Nota: Coeficientes no estandarizados de la regresión del criterio sobre la diversificación epistémica formativa (puntuaciones compuestas). En las pruebas por réplicas se informa la media y la desviación estándar de las réplicas; en las demás, el coeficiente y su error estándar. El predictor placebo conserva la distribución del predictor pero rompe su asociación con el criterio ($p = .102$).

Figura 1-2

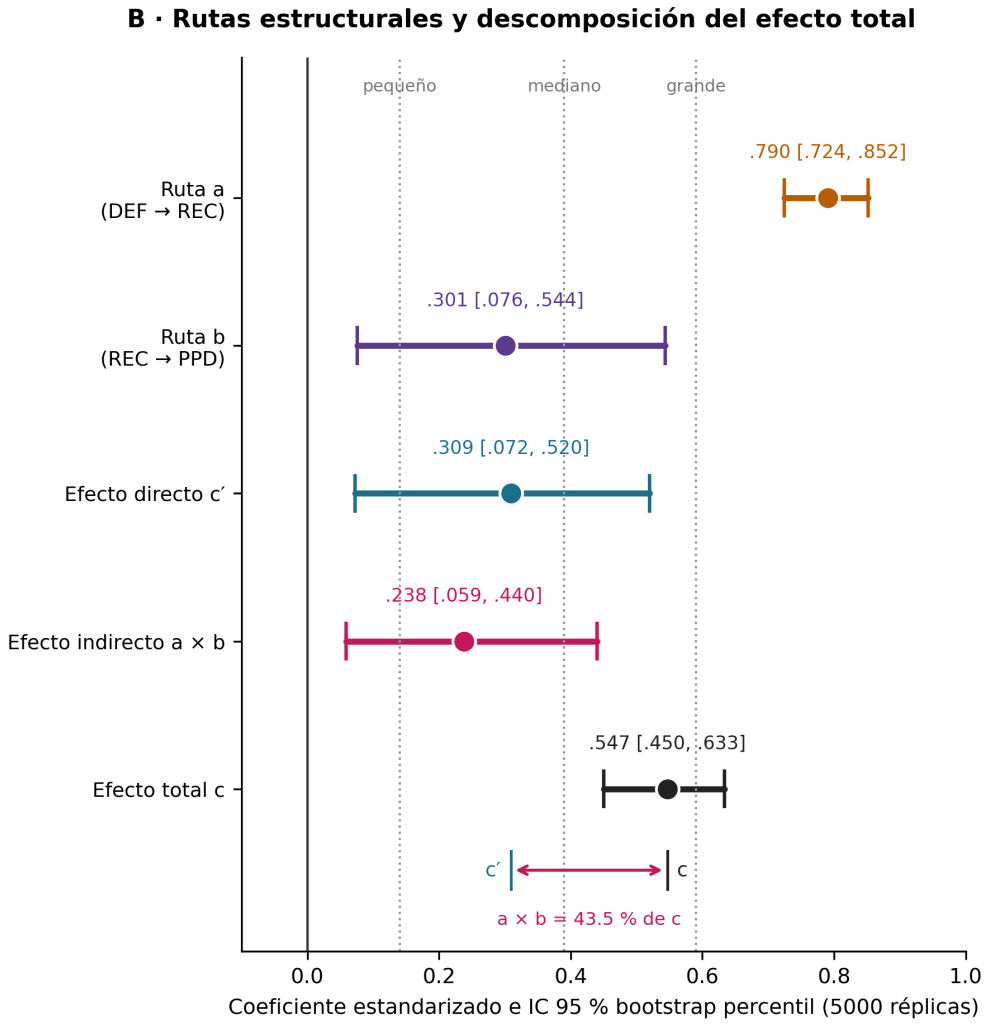
Modelo de ecuaciones estructurales completo y descomposición del efecto de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora.

A · Modelo de ecuaciones estructurales: medición y rutas estandarizadas

N = 375 · ML · $\chi^2(24) = 20.093$ · CFI = 1.000 · RMSEA = 0.000 · SRMR = .024



Nota. N = 375; CB-SEM estimado por máxima verosimilitud; coeficientes estandarizados con intervalos de confianza bootstrap percentil al 95 % (5000 réplicas, todas convergentes). Panel A: diagrama del modelo de medición y estructural; las elipses representan los constructos latentes y los rectángulos sus indicadores (DEF1-DEF3, REC1-REC3, PPD1-PPD3, en el orden de la Tabla 3), con su carga estandarizada sobre cada flecha. Rutas estructurales: a (H2) = .790, p < .001, IC 95 % [.724, .852]; b = .301, p = .014, IC 95 % [.076, .544]; c′ (H1) = .309, p = .008, IC 95 % [.072, .520]. Efecto indirecto (H3): a × b = .238, IC 95 % [.059, .440]; efecto total: c = .547, IC 95 % [.450, .633]. Ajuste global: $\chi^2(24) = 20.093$, CFI = 1.000, RMSEA = .000, SRMR = .024. *p < .05; **p < .01; ***p < .001.



Nota. N = 375; CB-SEM estimado por máxima verosimilitud; coeficientes estandarizados con intervalos de confianza bootstrap percentil al 95 % (5000 réplicas, todas convergentes). Panel B: estimación puntual e intervalo bootstrap de las rutas a y b y de los efectos directo, indirecto y total; las líneas punteadas verticales señalan los valores de referencia de efecto pequeño (.14), mediano (.39) y grande (.59) empleados por Fritz y MacKinnon (2007). La flecha inferior muestra que el efecto indirecto equivale al 43.5 % del efecto total (.238/.547) y el directo al 56.5 %; ambos intervalos excluyen el cero, lo que corresponde a una mediación parcial.

Tabla 1-9

Regresión jerárquica de la práctica pedagógica descolonizadora sobre puntuaciones compuestas (análisis de robustez)

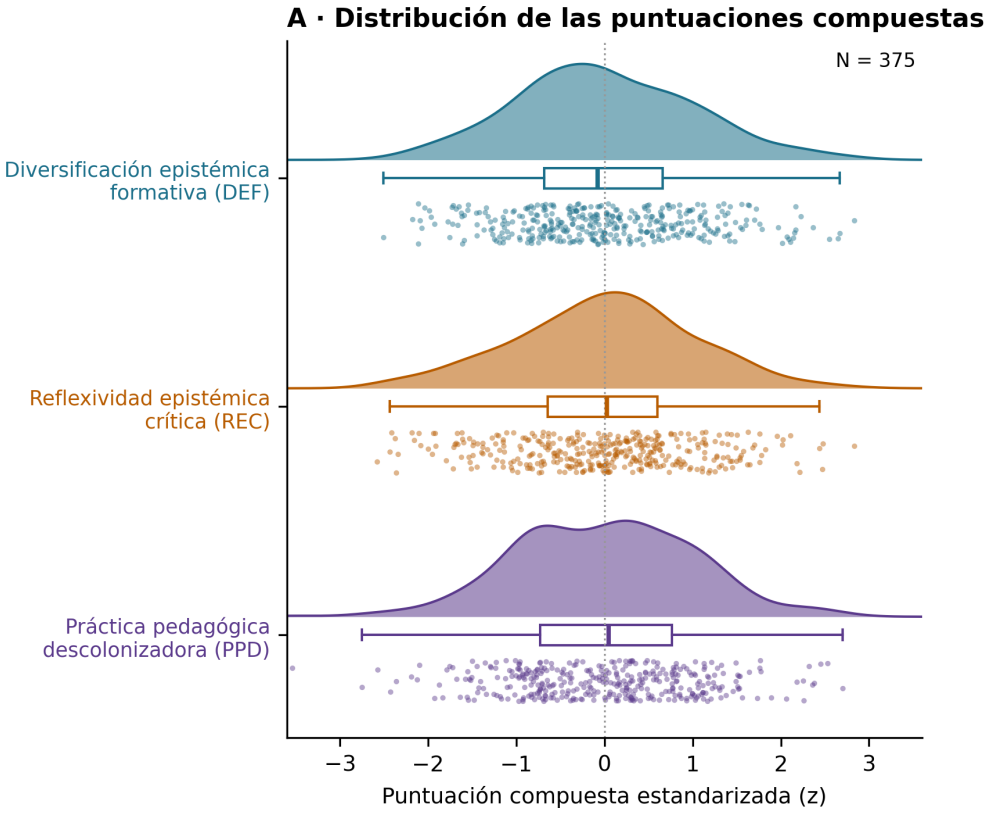
Modelo	Predictor	b	EE (HC3)	IC 95 %	t
1	Diversificación epistémica formativa	.242***	.020	[.203, .282]	11.879
2	Diversificación epistémica formativa	.145***	.030	[.087, .203]	4.879
3	Reflexividad epistémica crítica	.231***	.056	[.122, .341]	4.138

Modelo	Predictor	R^2			F
		base	ajustado	Δ	
1	Diversificación epistémica formativa	.278	.276	—	F(1, 373) = 141.109, p < .001
2	Diversificación epistémica formativa	.308	.304	.031	F(2, 372) = 82.633, p < .001
3	Reflexividad epistémica crítica				F de cambio(1, 372) = 16.51, p < .001

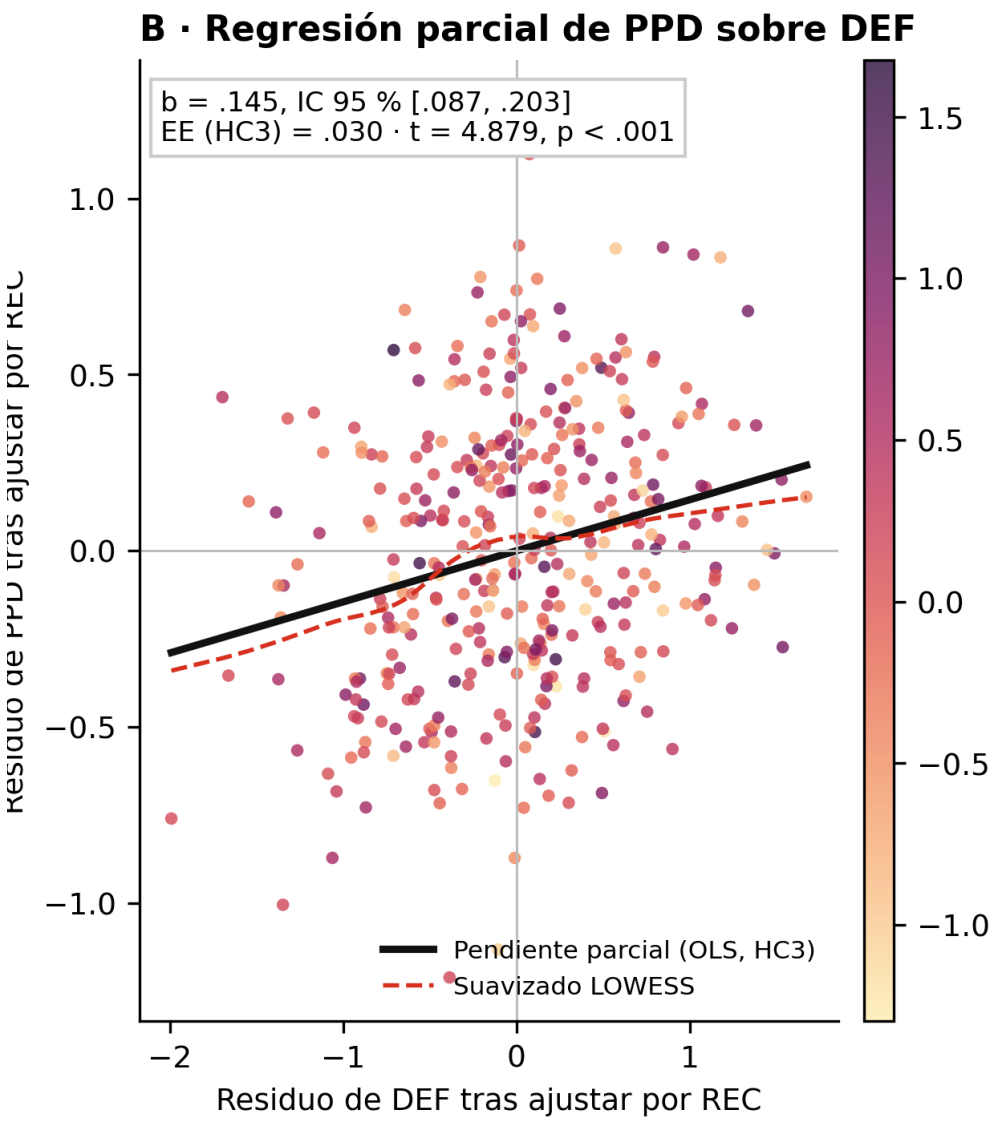
Nota: N = 375. Coeficientes no estandarizados sobre puntuaciones compuestas, con errores estándar robustos HC3. ***p < .001. Al no declararse moderadores en el modelo, la secuencia jerárquica se reduce a dos bloques. Los coeficientes no son comparables en magnitud con los β latentes de la Tabla 7; la tabla verifica la dirección y la significación de las rutas con otro estimador.

Figura 1-3

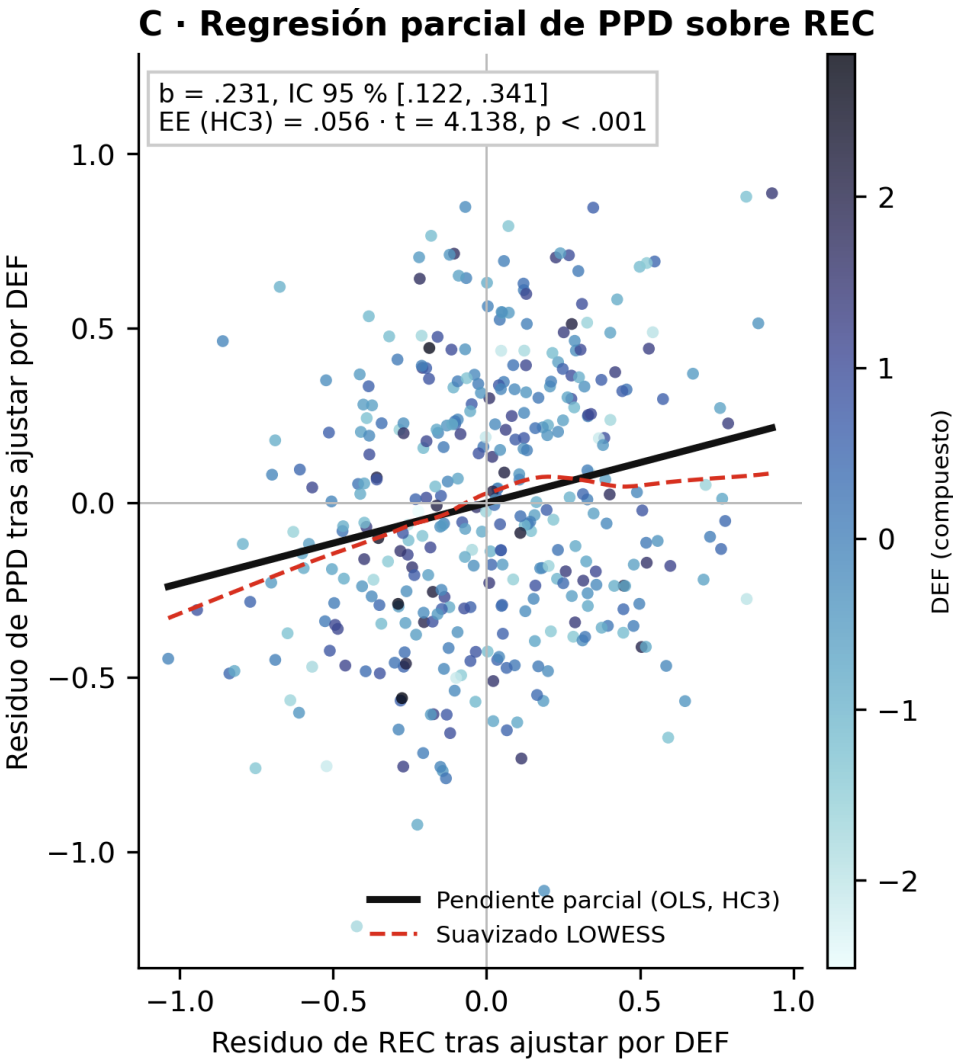
Distribución de las puntuaciones compuestas y regresión parcial de la práctica pedagógica descolonizadora



Nota. N = 375. Panel A: distribución de las puntuaciones compuestas de los tres constructos, estandarizadas (z) para su comparación; cada fila combina la densidad de kernel, un diagrama de caja (mediana y rango intercuartílico, bigotes hasta 1.5 veces dicho rango) y los casos individuales con desplazamiento vertical aleatorio para evitar superposición.



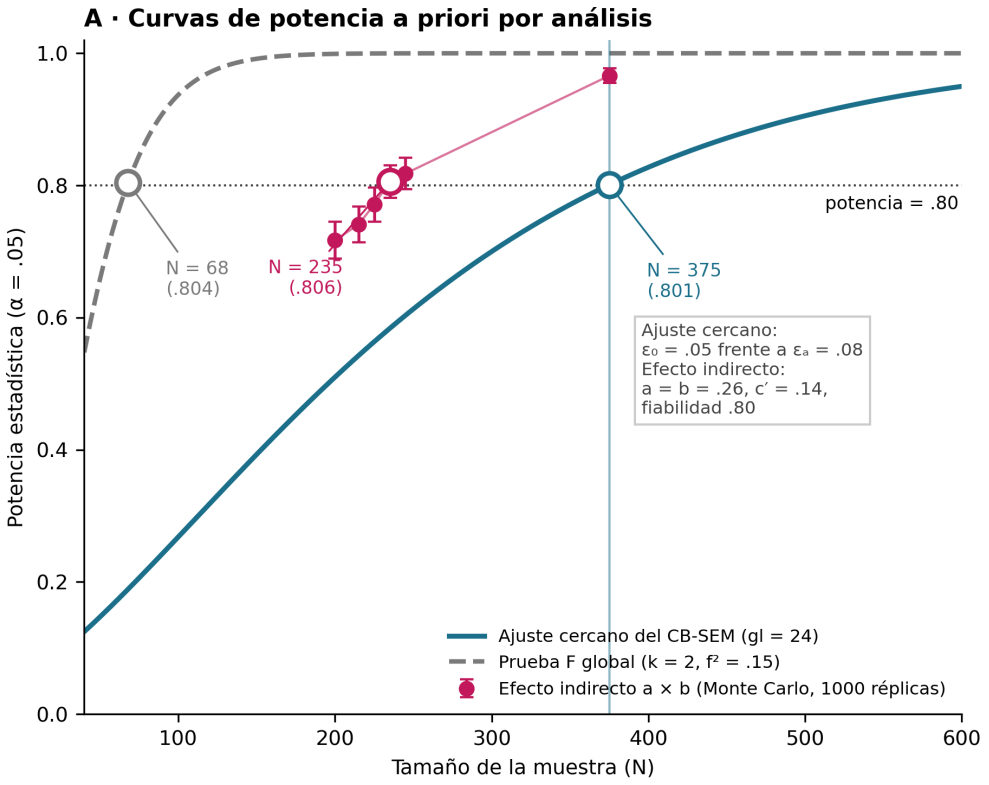
Nota. N = 375. Paneles B y C: gráficos de regresión parcial (variable añadida) del modelo de regresión jerárquica sobre puntuaciones compuestas, estimado por mínimos cuadrados ordinarios con errores estándar robustos HC3 (análisis de robustez, Tabla 1-9, Modelo 2; R^2 = .308); cada eje muestra los residuos tras ajustar por el otro predictor, de modo que la pendiente de la recta negra coincide con el coeficiente parcial. Panel B: residuo de PPD frente al residuo de DEF ajustando por REC, b = .145, EE = .030, IC 95 % [.087, .203], t = 4.879, p < .001; el color de cada punto indica su puntuación compuesta en REC



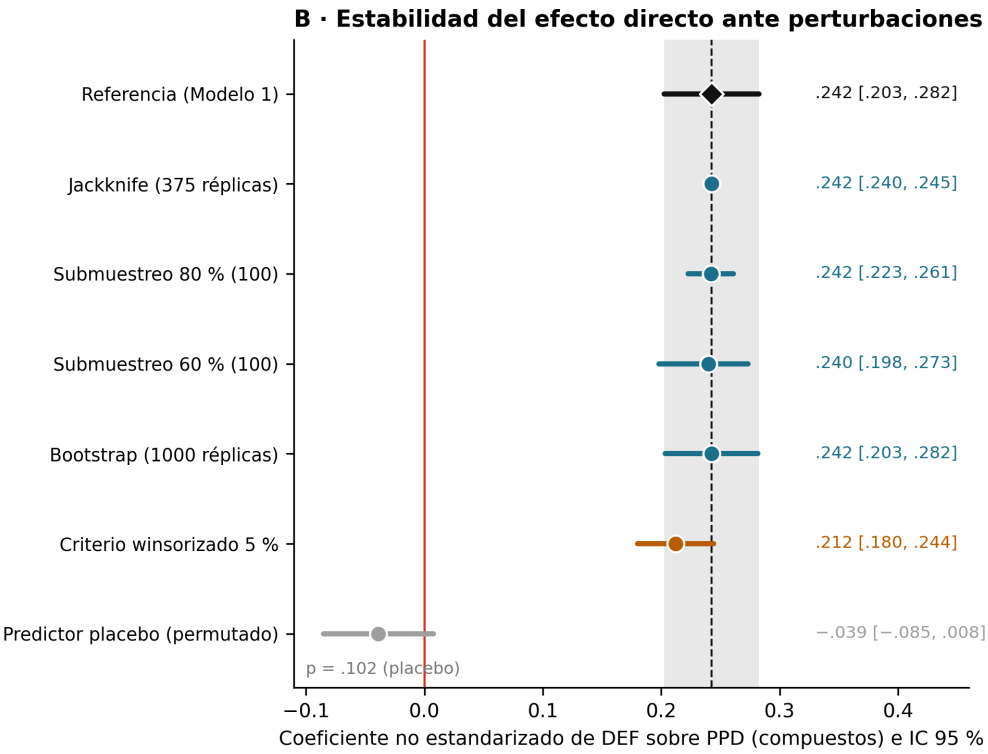
Nota. N = 375. Paneles B y C: gráficos de regresión parcial (variable añadida) del modelo de regresión jerárquica sobre puntuaciones compuestas, estimado por mínimos cuadrados ordinarios con errores estándar robustos HC3 (análisis de robustez, Tabla 1-9, Modelo 2; $R^2 = .308$); cada eje muestra los residuos tras ajustar por el otro predictor, de modo que la pendiente de la recta negra coincide con el coeficiente parcial. Panel C: residuo de PPD frente al residuo de REC ajustando por DEF, b = .231, EE = .056, IC 95 % [.122, .341], t = 4.138, p < .001; el color indica la puntuación compuesta en DEF. La línea roja discontinua es un suavizado LOWESS (fracción .60) que permite valorar la linealidad de la relación. Los coeficientes son no estandarizados y proceden de un estimador distinto del modelo latente de la Figura 2, por lo que no son comparables en magnitud con sus β.

Figura 1-4

Potencia a priori por análisis y estabilidad del efecto directo



Nota. Panel A: potencia estadística en función del tamaño de la muestra (α = .05) para los tres análisis del estudio. La curva continua corresponde a la prueba de ajuste cercano del CB-SEM (MacCallum et al., 1996; gl = 24, ε_0 = .05 frente a ε_a = .08), calculada con la distribución chi-cuadrado no central; alcanza .80 con N = 375 (potencia de .801). La curva discontinua corresponde a la prueba F global de la regresión del criterio sobre dos predictores (f^2 = .15; Hair et al., 2022), que alcanza .80 con N = 68 (.804). Los puntos corresponden a la simulación Monte Carlo del efecto indirecto (1000 réplicas por tamaño, intervalos bootstrap percentil al 95 % con 1000 remuestreos; a = b = .26, c′ = .14, fiabilidad .80), con barras de ± 1.96 errores estándar binomiales; la potencia llega a .806 con N = 235 y a .966 con N = 375. La línea vertical marca el tamaño de la muestra del estudio (N = 375), fijado por el análisis más exigente (Tabla 1-2).



Nota. Panel B: coeficiente no estandarizado de la diversificación epistémica formativa sobre la práctica pedagógica descolonizadora (regresión sobre puntuaciones compuestas) e IC 95 % en la estimación de referencia (b = .242 [.203, .282]; banda gris y línea discontinua) y bajo cinco perturbaciones de la muestra: jackknife con exclusión de un caso (375 réplicas), submuestreos del 80 % y del 60 % (100 réplicas cada uno) y bootstrap (1000 réplicas), en los que se informa la media de las réplicas y su intervalo, todos con el signo y la significación conservados en el 100 % de las réplicas; y winsorización del criterio al 5 % (b = .212 [.180, .244]). Con el predictor placebo obtenido por permutación el coeficiente se anula (b = −.039 [−.085, .008], p = .102); la línea roja marca el cero (Tabla 1-10).

El efecto directo de la diversificación epistémica formativa en la regresión sobre compuestos conservó el signo y la significación en todas las réplicas jackknife, en los submuestreos del 80 % y del 60 % y en las réplicas bootstrap, así como al winsorizar el criterio al 5 % (Tabla 1-10).

Con el predictor placebo permutado, el coeficiente se anuló. La Figura 1-4 resume la potencia a priori de los tres análisis y esta prueba de estabilidad.

1.5 DISCUSIÓN

El hallazgo más desconcertante no es la mediación esperada, sino que el constructo llamado a transmitir el efecto resulta casi indistinguible de

su antecedente: la ruta de la diversificación hacia la reflexividad es muy intensa, la correlación entre ambos constructos supera la raíz de la AVE de la reflexividad y, pese a ello, el efecto directo sobre la práctica no se disuelve, sino que sigue siendo mayor que el indirecto.

Que una reflexividad tan próxima a la formación absorba menos de la mitad de la asociación obliga a una lectura cautelosa. La evidencia de simulación según la cual el criterio de Fornell y Larcker no detecta de forma fiable la falta de validez discriminante, y el mejor desempeño del HTMT (Henseler et al., 2015), respalda que ambos constructos son separables, aunque la proximidad sugiere que, en la experiencia de estos docentes, haber sido formados en un currículo plural y examinar críticamente el propio lugar epistémico son facetas contiguas de un mismo proceso. Con esa salvedad, los resultados respaldan H1, H2 y H3 y son consistentes con la idea de que la descolonización de la pedagogía no se agota en la inclusión de contenidos, sino que requiere que el docente interrogue la posición desde la que enseña (Allen y Girei, 2024; R'boul et al., 2026).

La magnitud de la ruta que sostiene H2 es coherente con la evidencia de que las pedagogías relacionales, al permitir a los futuros docentes conectar con formas de conocer indígenas, aumentan su disposición a incorporarlas (Macdonald et al., 2024), y con la caracterización de la reflexividad como instancia que media entre las condiciones de la formación y la agencia profesional (Hinostroza, 2023). Asimismo, la noción de meta-reflexividad como componente del profesionalismo docente (Cramer et al., 2023) ofrece un lenguaje afín para el mecanismo de conversión aquí estimado.

Más matizada es la relación con la literatura sobre la distancia entre disposiciones y práctica. En una muestra de 606 docentes de escuelas culturalmente diversas, las creencias sobre la diversidad apenas se correlacionaron con las prácticas culturalmente relevantes (Liu et al., 2021), mientras que aquí la asociación entre formación y práctica fue moderada y se transmitió, en parte, por un mecanismo identificable. Esta divergencia admite varias lecturas que los datos no permiten dirimir. Una es que la variable formativa, definida como una experiencia epistémica

concreta y no como una creencia general, esté más cerca de la conducta que las creencias; otra, que las medidas de autorreporte de una misma fuente inflen la covariación entre formación y práctica; y una tercera, que el contexto chino y el peruano difieran en aspectos que no se midieron. De manera análoga, en 417 docentes de secundaria, la autoeficacia y la formación profesional orientada a la equidad se relacionaron con la enseñanza culturalmente responsiva (Comstock et al., 2023), lo que sugiere que el contenido de la formación importa, pero no autoriza a afirmar que sea su contenido epistémico, y no su mera existencia, lo decisivo, pues el diseño no compara tipos de formación. Desde una perspectiva crítica, tampoco cabe descartar que la persistencia del efecto directo refleje repertorios de actividades y materiales que los docentes reproducen sin un examen explícito de sus supuestos, un proceso que la formación culturalmente responsiva con jóvenes refugiados, centrada en prácticas de colaboración de docentes en formación, sugiere como vía transformadora distinta de la introspección (Walker, 2023). Si esto es así, una escala de tres ítems de reflexividad podría captar solo una parte del examen epistémico, y la fracción mediada estaría subestimada.

En el plano de la política educativa, los resultados ofrecen una explicación tentativa de por qué los currículos que declaran la interculturalidad no siempre la producen en el aula. El Perú ha institucionalizado la EIB, pero la diversidad lingüística y cultural sigue ocupando espacios marginales en la escuela (Flores et al., 2023; Sullón, 2024), y la implementación del nuevo currículo de formación docente EIB enfrenta limitaciones de colaboración y rotación de formadores (Moreano et al., 2026). En ese escenario, diversificar el currículo formativo sería necesario, pero es consistente con los datos que no baste si no se acompaña de espacios deliberados de análisis de las jerarquías del conocimiento que los propios docentes reproducen. La literatura sobre la formación docente en el país apunta en la misma dirección desde ángulos diversos: los dilemas de una formación (de)colonial en los Andes (Zavala, 2025), los reconocimientos identitarios y las trayectorias de revitalización lingüística en la formación docente EIB, que subrayan su dimensión reflexiva (Kvietok, 2025), la formación

intercultural bilingüe en comunidades nativas de Amazonas (Chávez et al., 2026) y la comunidad educativa indígena de Formabiap, que opera como pedagogía alternativa al canon (Rahman et al., 2023). Una revisión de alcance sobre identidad cultural y descolonización en escuelas rurales (Voto, 2026) sitúa estas cuestiones en el nivel de la práctica pedagógica que aquí se modeló.

La discusión se extiende a la región, donde las instituciones formadoras son el escenario principal. El análisis de la descolonialidad y la interculturalidad crítica en la educación superior de la Amazonía ecuatoriana (Arias y Minoia, 2023) y del papel de las universidades de inspiración indígena (Guzmán et al., 2025) permite pensar la diversificación como una transformación institucional y no solo curricular. La propuesta de formar investigadores desde la epistemología del Sur en la formación inicial docente (Parada et al., 2024) sería una forma concreta de cultivar la reflexividad que el modelo identifica como mecanismo, y la pregunta por qué interculturalidad promueve la política de educación superior chilena (Vazquez, 2026) recuerda que no toda diversificación declarada equivale a diversificación epistémica.

El argumento, por último, podría trascender el caso peruano. En sistemas tan distintos como el finlandés o el chileno, la distancia entre el discurso institucional y el conocimiento efectivamente transmitido sobre los pueblos indígenas parece responder a una lógica semejante (Bustos et al., 2026; Mattila et al., 2025), lo que convierte el mecanismo identificado en una hipótesis de alcance general para la formación docente en contextos poscoloniales, no en un resultado generalizable por sí mismo.

1.5.1 Limitaciones y líneas futuras

El estudio presenta limitaciones que acotan la interpretación de sus resultados. Primero, el diseño transversal impide establecer precedencia temporal entre los constructos, por lo que el efecto indirecto debe leerse como una descomposición de asociaciones compatible con el mecanismo propuesto y no como evidencia causal; un diseño longitudinal con

mediciones durante y después de la formación permitiría contrastar la secuencia. Segundo, el muestreo por conveniencia limita la generalización a la población docente del país. Tercero, todas las medidas son de autorreporte y proceden de una misma fuente, lo que expone los resultados a varianza común de método; la incorporación de observaciones de aula o de análisis documental de planificaciones fortalecería la medición de la práctica. Cuarto, la cercanía entre la diversificación epistémica formativa y la reflexividad epistémica crítica aconseja refinar las escalas con más ítems y con contenidos que diferencien mejor la experiencia formativa recibida de la disposición reflexiva del docente. Quinto, no se evaluó la invarianza de la medición entre grupos de docentes (por ejemplo, según región, tipo de institución formadora o participación en programas de EIB); contrastarla permitiría examinar si el mecanismo opera de forma equivalente en los distintos contextos del país.

1.6 CONCLUSIONES

La diversificación epistémica de la formación docente se asocia con la práctica pedagógica descolonizadora de los docentes de humanidades, de manera directa y a través de la reflexividad epistémica crítica. Esta última actúa como un mecanismo de conversión que explica el 43.5 % de la asociación total entre lo que la formación ofrece y lo que el docente hace en el aula, de modo que la mediación es parcial y coexiste con una vía directa que el modelo no desagrega. Las tres hipótesis resultaron soportadas, y la dirección de los resultados se mantuvo con un estimador alternativo y frente a perturbaciones de la muestra.

Para la teoría de la pedagogía descolonial, el estudio aporta la especificación y el contraste empírico de ese mecanismo, que hasta ahora se había formulado sobre todo en términos conceptuales. Para la investigación educativa, muestra que constructos de raíz descolonial pueden medirse con fiabilidad y validez aceptables y modelarse mediante ecuaciones estructurales basadas en covarianzas. Para la política de formación docente, sugiere que diversificar el currículo formativo es necesario, pero

probablemente insuficiente si no se acompaña del cultivo deliberado de la reflexividad epistémica, por ejemplo mediante espacios de análisis de las jerarquías del conocimiento que los propios docentes reproducen.

Estas conclusiones deben leerse con la cautela que imponen un diseño transversal, una muestra por conveniencia, medidas de autorreporte y la cercanía empírica entre formación y reflexividad. La agenda inmediata consiste en replicar el modelo con diseños longitudinales, incorporar observación de aula, refinar la medición de la reflexividad y examinar la invarianza entre contextos formativos, de modo que el mecanismo propuesto pueda someterse a pruebas más exigentes en sistemas educativos que declaran la interculturalidad en sus currículos.

1.7 REFERENCIAS BIBLIOGRÁFICAS

Akbar, A. (2024). On the epistemic urgency of decolonizing the school curriculum: a reflection. *Journal of Philosophy of Education, 58*(2-3), 397-411. https://doi.org/10.1093/jopedu/qhad081

Allen, S., & Girei, E. (2024). Developing Decolonial Reflexivity: Decolonizing Management Education by Confronting White Skin, White Identities, and Whiteness. *Academy of Management Learning & Education, 23*(2), 246-264. https://doi.org/10.5465/amle.2022.0387

Arias, R., & Minoia, P. (2023). Decoloniality and Critical Interculturality in Higher Education: Experiences and Challenges in Ecuadorian Amazonia. *Forum for Development Studies, 50*(1), 11-34. https://doi.org/10.1080/08039410.2023.2177562

Bustos, R., Mondaca, C., Driscoll, T., & Sanchez, E. (2026). Decolonization and curriculum: Representations of Chilean indigenous teachers. *Teaching and Teacher Education, 174,* 105413. https://doi.org/10.1016/j.tate.2026.105413

Chávez, R., Ruiz, Y., Carranza, R., Campos, J., Guevara, R., Navas, C., & Poémape, C. (2026). Intercultural bilingual education in teacher education within native communities of the Amazonas region. *Studies in Higher Education*, 1-20. https://doi.org/10.1080/03075079.2026.2679171

Comstock, M., Litke, E., Hill, K., & Desimone, L. (2023). A Culturally Responsive Disposition: How Professional Learning and Teachers' Beliefs About and Self-Efficacy for Culturally Responsive Teaching Relate to Instruction. *AERA Open, 9*. https://doi.org/10.1177/23328584221140092

Cramer, C., Brown, C., & Aldridge, D. (2023). Meta-Reflexivity and Teacher Professionalism: Facilitating Multiparadigmatic Teacher Education to Achieve a Future-Proof Profession. *Journal of Teacher Education, 74*(5), 467-480. https://doi.org/10.1177/00224871231162295

Flores, H., Gamarra, J., Mamani, C., & Vilcarromero, L. (2023). Intercultural bilingual education (EIB) in Peru. *Research in Subject-matter Teaching and Learning, 6*(1), 84-95. https://doi.org/10.2478/ristal-2023-0006

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Fritz, M., & MacKinnon, D. (2007). Required Sample Size to Detect the Mediated Effect. *Psychological Science, 18*(3), 233-239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Gandarilla, J., García, M., & Benzi, D. (2021). Two Decades of Aníbal Quijano's Coloniality of Power, Eurocentrism and Latin America. *Contexto Internacional*, *43*(1), 199-222. https://doi.org/10.1590/s0102-8529.2019430100009

Guzmán, C., Chiappa, R., & Gómez, C. (2025). Transforming regional higher education: the decolonising role of Indigenous-inspired universities in Latin America. *International Journal of Educational Research*, *133*, 102731. https://doi.org/10.1016/j.ijer.2025.102731

Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.

Hayes, A. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). Guilford Press.

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hinostroza, Y. (2023). Reflexivity and agency in university-based teacher educators: a critical realist analysis. *Teaching in Higher Education*, *28*(8), 1918-1936. https://doi.org/10.1080/13562517.2021.1943651

Kline, R. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Kvietok, F. (2025). Reconocimientos identitarios y trayectorias de revitalización lingüística en la formación docente en educación

intercultural bilingüe. *Íkala, Revista de Lenguaje y Cultura, 30*(3). https://doi.org/10.17533/udea.ikala.360321

L'Estrange, L., Walker, S., Lunn, J., Ryan, M., Bourke, T., Rowan, L., & Johansson, E. (2023). Capturing epistemic reflexivity for teacher educators teaching about/to/for diversity in teacher education. *Heliyon, 9*(2), e13754. https://doi.org/10.1016/j.heliyon.2023.e13754

Liu, Q., Çolak, F., & Agirdag, O. (2021). Teachers' beliefs and practices in culturally diverse schools: an empirical study in Southwest China. *Teachers and Teaching, 27*(7), 672-695. https://doi.org/10.1080/13540602.2021.1987878

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Macdonald, M., Booth, S., & Jackson, L. (2024). An ecosystem of knowledge: relationality as a framework for teachers to infuse Indigenous perspectives in curriculum. *Asia-Pacific Journal of Teacher Education, 52*(2), 175-192. https://doi.org/10.1080/1359866x.2024.2314285

Mattila, E., Lindén, J., & Annala, J. (2025). On the shoulders of a perfect stranger: knowledge gap about the Indigenous Sámi in the Finnish teacher education curriculum. *Race Ethnicity and Education, 28*(7), 1157-1173. https://doi.org/10.1080/13613324.2023.2249292

Meshcheryakov, G., Igolkina, A., & Samsonova, M. (2021). semopy 2: A structural equation modeling package with random effects in Python. *arXiv*. https://doi.org/10.48550/arXiv.2106.01140

Moreano, G., Zegarra, C., & Rivera, Y. (2026). New intercultural bilingual education teacher preparation curriculum: Opportunities and challenges to reduce educational inequities in Peru. *Education Policy Analysis Archives, 34*. https://doi.org/10.14507/epaa.34.9008

Parada, M., Lárez, J., & Vega, Ó. (2024). Building knowledge from the epistemology of the South: the importance of training researchers in initial teacher training. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1231602

Quijano, A. (2000). Coloniality of Power and Eurocentrism in Latin America. *International Sociology*, *15*(2), 215-232. https://doi.org/10.1177/0268580900015002005

R'boul, H., Boulahnane, S., & Benzizoune, O. (2026). Decentring humanities education through intercultural studies: the decolonial gesture and epistemic authority. *Teaching in Higher Education*, 1-19. https://doi.org/10.1080/13562517.2026.2746686

Rahman, E., Barbira, F., García, F., & Castro, M. (2023). Formabiap's Indigenous educative community, Peru: a biosocial pedagogy. *Oxford Review of Education*, *49*(4), 536-554. https://doi.org/10.1080/03054985.2023.2223921

Santos, B. (2014). *Epistemologies of the South: Justice against epistemicide*. Routledge. https://doi.org/10.4324/9781315634876

Stopford, R. (2025). Decolonising the curriculum: common sense, threshold concepts, and epistemic injustice. *Higher Education, 90*(3), 861-879. https://doi.org/10.1007/s10734-024-01354-3

Sullón, K. (2024). El paisaje lingüístico en las escuelas de Educación Intercultural Bilingüe del Perú. *Revista Iberoamericana de Educación, 96*(1), 83-95. https://doi.org/10.35362/rie9616436

Thakur, D., & Mukherjee, S. (2026). Decolonizing Internationalization: A Southern Epistemic Framework for Higher Education Reform. *Higher Education Policy, 39*(3), 962-987. https://doi.org/10.1057/s41307-025-00409-z

Vazquez, L. (2026). Which interculturality? Intercultural bilingual education and intercultural competence in Chilean higher education policy. *Intercultural Education,* 1-21. https://doi.org/10.1080/14675986.2026.2712121

Voto, M. (2026). Identidad cultural y descolonización educativa en escuelas rurales del Perú: una revisión de alcance. *Revista Enfoques Educacionales, 23*(2), 335-357. https://doi.org/10.5354/2735-7279.2026.83287

Walker, A. (2023). Transformative Potential of Culturally Responsive Teaching: Examining Preservice Teachers' Collaboration Practices Centering Refugee Youth. *Education Sciences, 13*(6), 621. https://doi.org/10.3390/educsci13060621

Zavala, V. (2025). Disentangling (de)colonial teacher education in the Andes: Dilemmas and possibilities. *The Modern Language Journal, 109*(3), 550-568. https://doi.org/10.1111/modl.13012

Zembylas, M. (2025). Beyond teacher impartiality: addressing epistemic-affective injustice in times of political polarization. *Ethics and Education, 20*(4), 424-437. https://doi.org/10.1080/17449642.2025.2565034

CAPÍTULO II

ÉTICA BIOCÉNTRICA, INTEGRACIÓN CURRICULAR ECOPEDAGÓGICA Y CONDUCTA PROAMBIENTAL DOCENTE: UN MODELO DE ECUACIONES ESTRUCTURALES CON MEDIACIÓN PARCIAL EN LA EDUCACIÓN BÁSICA DEL PERÚ

Wilder Antonio Vega Alcántara

Universidad Nacional del Centro del Perú

Resumen

La educación ambiental exige hoy formar disposiciones que reconozcan el valor de la vida no humana, y el docente es el agente que traduce sus valores en decisiones curriculares y en conducta. Se desconoce, sin embargo, qué fracción del vínculo entre ética y conducta transita por la práctica curricular. Este estudio cuantitativo, no experimental y transversal estimó, mediante un modelo de ecuaciones estructurales basado en covarianzas, la relación entre ética biocéntrica, integración curricular ecopedagógica y conducta proambiental en 375 docentes de educación básica del Perú, seleccionados por conveniencia. El modelo de medición mostró cargas entre 0,686 y 0,845, fiabilidad compuesta entre 0,796 y 0,837 y validez discriminante satisfactoria. El modelo estructural presentó un ajuste excelente (CFI = 0,997; RMSEA = 0,020) y superó al modelo de un factor. La ética biocéntrica se asoció con la integración curricular (β = 0,518), esta con la conducta (β = 0,397) y la ética conservó una asociación directa con la conducta (β = 0,420). El efecto indirecto fue de 0,206 (IC95 % [0,131; 0,286]) y representó el 32,9 % del efecto total, de modo que la mediación fue parcial. Los resultados sugieren que la formación docente debe atender tanto las creencias éticas como la competencia de planificación ecopedagógica. El diseño transversal, el muestreo no probabilístico y el autorreporte limitan la inferencia causal y la generalización.

Palabras clave: ética biocéntrica; ecopedagogía; integración curricular; conducta proambiental; docentes; ecuaciones estructurales

Abstract

Environmental education now requires cultivating dispositions that recognize the value of non-human life, and teachers are the agents who translate their values into curricular decisions and behavior. Yet it remains unknown what share of the link between ethics and behavior runs through curricular practice. This quantitative, non-experimental, cross-sectional study used covariance-based structural equation modeling to estimate the relationships among biocentric

ethics, ecopedagogical curricular integration, and pro-environmental behavior in 375 Peruvian basic-education teachers selected by convenience. The measurement model showed loadings between 0.686 and 0.845, composite reliability between 0.796 and 0.837, and satisfactory discriminant validity. The structural model showed excellent fit (CFI = 0.997; RMSEA = 0.020) and outperformed a one-factor model. Biocentric ethics was associated with curricular integration (β = 0.518), curricular integration with behavior (β = 0.397), and ethics retained a direct association with behavior (β = 0.420). The indirect effect was 0.206 (95% CI [0.131, 0.286]) and accounted for 32.9% of the total effect, indicating partial mediation. The results suggest that teacher education should address both ethical beliefs and the competence for ecopedagogical planning. The cross-sectional design, non-probability sampling, and self-report measures limit causal inference and generalization.

Keywords: *biocentric ethics; ecopedagogy; curricular integration; pro-environmental behavior; teachers; structural equation modeling*

2.1 INTRODUCCIÓN

La crisis ecológica planetaria ha desplazado la educación ambiental desde un contenido temático más hacia una exigencia ética y curricular de la escuela: los sistemas educativos ya no solo deben informar sobre el deterioro de los ecosistemas, sino formar disposiciones y prácticas que reconozcan el valor de la vida no humana (UNESCO, 2021). La revisión bibliométrica de los Objetivos de Desarrollo Sostenible y la educación en América Latina da cuenta de la amplitud y la dispersión de ese desafío regional (Pedraja et al., 2023), y reformas curriculares recientes, como la ecuatoriana de 2024, muestran que los valores se han convertido en objeto explícito de política educativa, aunque su recepción pública se haya encuadrado sobre todo como un correctivo moral frente a la erosión de valores (Tusa et al., 2025). En esa transición, el docente ocupa una posición bisagra, porque sus valores orientan qué decide enseñar, cómo lo organiza en el currículo y qué conductas modela ante sus estudiantes.

El fundamento normativo de esa exigencia procede de la ética ambiental. El debate entre las posiciones antropocéntricas y biocéntricas, que difieren en el estatuto moral que atribuyen a los seres vivos, ha sido sistematizado recientemente en sus bases teóricas (Ashurov, 2026), y la ética biocéntrica sostiene el reconocimiento del valor intrínseco de todo ser vivo con independencia de su utilidad para el ser humano (Taylor, 1986). Esta tradición se ha prolongado en propuestas contemporáneas que fundamentan una ética global a partir de la rareza y la fragilidad de la biosfera terrestre (Gleiser, 2025), que examinan las bases de una igualdad biocéntrica (McShane, 2024) o que sitúan la responsabilidad biocéntrica en el tránsito hacia la justicia ambiental (Barzola y Agoglia, 2024). Su traducción educativa es la crítica a los marcos escolares que refuerzan implícitamente la supremacía humana, a favor de una solidaridad interespecie y de un aprendizaje relacional y encarnado (McIntosh y Feltrin, 2025).

La ecopedagogía ha formulado esa traducción como una reorientación ética del acto educativo hacia la relación entre sociedad y naturaleza (Boxley, 2024), y la ha ampliado hacia una educación para los co-seres (Hung, 2025). Sus fundamentos y alcances, no exentos de debate, han sido delimitados en la literatura de habla hispana (Ruiz et al., 2021). De ese marco se deriva el constructo que interesa aquí: la integración curricular ecopedagógica, entendida como la incorporación que el docente efectivamente planifica y ejecuta de contenidos, proyectos y saberes locales con orientación ecopedagógica.

Un problema central para la formación docente permanece, no obstante, sin resolver: si la ética del profesorado se traduce directamente en conducta proambiental o lo hace a través de su expresión pedagógica. La teoría valor-creencia-norma sostiene que las orientaciones de valor biosféricas activan creencias y normas personales que predisponen a la acción (Stern et al., 1999), pero no precisa cuánto de esa cadena se canaliza, en el caso del profesorado, por el currículo. La distinción no es menor. Si el efecto es fundamentalmente directo, bastaría con intervenir sobre las creencias; si una parte sustantiva transita por el currículo, las políticas

deben actuar también sobre la competencia de planificación y sobre las condiciones institucionales que la hacen posible. Los tres constructos suelen examinarse por separado y con frecuencia sin modelos que corrijan el error de medida, de modo que la magnitud relativa de las vías directa e indirecta permanece sin estimar.

El sistema educativo peruano ofrece un sitio de prueba especialmente informativo para esta cuestión. El Currículo Nacional de la Educación Básica incorpora un enfoque ambiental transversal que exige al docente traducir valores en experiencias de aprendizaje, en el marco de las políticas ambientales comparadas de la región (Gallegos y Rhoades, 2026), y lo hace en un territorio de alta biodiversidad y elevada presión sobre sus ecosistemas, donde la educación ambiental y la sostenibilidad institucional ya han sido descritas en instituciones públicas de regiones como Amazonas (Chávez, 2023).

La pregunta que orienta el estudio es: ¿en qué medida la ética biocéntrica de los docentes de educación básica se asocia con su conducta proambiental, de manera directa y a través de la integración curricular ecopedagógica? Se contrastan cuatro hipótesis: **H1**: la ética biocéntrica se asocia positivamente con la integración curricular ecopedagógica; **H2**: la integración curricular ecopedagógica se asocia positivamente con la conducta proambiental docente; **H3**: la ética biocéntrica se asocia positivamente con la conducta proambiental docente, controlando la integración curricular; **H4**: la ética biocéntrica ejerce un efecto indirecto positivo sobre la conducta proambiental docente a través de la integración curricular ecopedagógica.

La contribución es teórica y metodológica de alcance general: se estima, con un modelo de ecuaciones estructurales que corrige el error de medida, la fracción del vínculo entre ética y conducta que transita por la práctica curricular, un parámetro que la teoría ecopedagógica postula pero que no había sido cuantificado en el profesorado.

2.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

El marco principal es la teoría valor-creencia-norma (Stern et al., 1999), según la cual las orientaciones de valor, en particular los valores biosféricos, activan creencias y normas personales que predisponen a la acción proambiental. En este estudio, la ética biocéntrica operacionaliza el polo biosférico de esa cadena: un compromiso normativo con el valor intrínseco de la vida (Taylor, 1986). La teoría, no obstante, describe una secuencia general y deja abierta la cuestión de los canales específicos por los que los valores se convierten en acción en un rol profesional concreto, como el de quien enseña.

Como marco complementario se adopta la ecopedagogía crítica, que concibe el currículo como el espacio donde los valores ecológicos del docente se convierten en experiencias formativas situadas (Ruiz et al., 2021; Hung, 2025). Desde esta perspectiva, el currículo no es un mero vehículo de contenidos, sino la instancia en la que el docente decide qué merece ser enseñado y con qué sentido ético. La evidencia sobre el profesorado aporta, además, el principio de que las disposiciones valorativas se expresan en conducta a través de procesos intermedios más próximos a la acción, como la identidad ambiental y la práctica docente (Felix y Clayton, 2025; Öztürk y Pizmony, 2024). La conjunción de ambos marcos permite anticipar dos canales: uno normativo, en el que el valor opera como norma personal, y otro pedagógico, en el que opera mediante la planificación.

De estos marcos se derivan las hipótesis. Si la ética biocéntrica define lo que el docente considera valioso enseñar, debería asociarse con una mayor integración de contenidos y proyectos ecopedagógicos en su planificación (H1). Si esa integración estructura oportunidades recurrentes de acción ambiental en el aula y en la institución, debería asociarse con una conducta proambiental docente más frecuente (H2). Si la ética biocéntrica opera también como norma personal que guía la acción fuera del currículo, debería conservar una asociación directa con la conducta (H3). Y si la integración curricular es el mecanismo pedagógico que traduce valores

en práctica, la ética biocéntrica debería ejercer un efecto indirecto sobre la conducta a través de ella (H4).

El modelo resultante es de mediación parcial: se espera que coexistan una vía directa normativa y una vía indirecta pedagógica, y que el estudio pueda estimar el peso relativo de cada una.

2.3 MATERIALES Y MÉTODOS

2.3.1 Diseño

Estudio cuantitativo, no experimental, explicativo y de corte transversal, con recolección en 2025. La unidad de análisis fue el docente; aunque los participantes pertenecían a instituciones distintas, el análisis se realizó en un solo nivel.

2.3.2 Participantes y tamaño muestral.

Participaron 375 docentes en ejercicio de educación básica (inicial, primaria y secundaria) del Perú. La cifra corresponde al mínimo que asegura una potencia de 0,80 en la prueba de ajuste aproximado del modelo, el análisis más exigente, según el procedimiento de MacCallum et al. (1996) aplicado a 24 grados de libertad, contrastando $\varepsilon_0 = 0{,}05$ frente a $\varepsilon_1 = 0{,}08$ con $\alpha = 0{,}05$ mediante la distribución chi-cuadrado no central. Para la ecuación de la conducta, con dos predictores, se empleó la prueba F con $f^2 = 0{,}15$ y 0,10 (Cohen, 1988), y para el efecto indirecto una simulación de Monte Carlo con 2 000 réplicas y prueba de significación conjunta, fijando $a = b = 0{,}26$ y $c' = 0{,}10$ (Tabla 2-2). Estas magnitudes se anclaron de forma conservadora: quedan por debajo del umbral convencional de efecto moderado y de las asociaciones reportadas con docentes peruanos ($r = 0{,}523$–$0{,}603$, Carlos et al., 2025; incidencia del 21 %, Vásquez et al., 2022); la menor de esas correlaciones equivale a $f^2 = 0{,}377$. La admisibilidad del modelo se verificó por álgebra, con nueve indicadores, $N \geq 10$ y 24 grados de libertad. La selección fue no probabilística, por conveniencia, mediante invitación a docentes de instituciones públicas y privadas accesibles al equipo; por ello

no se afirma representatividad. Se incluyeron docentes con al menos un año de experiencia y se excluyeron los cuestionarios incompletos.

2.3.3 Instrumentos

Los tres constructos se midieron como variables latentes reflectivas con escalas Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo), de tres ítems cada una (Tabla 2-1).

Tabla 2-1

Operacionalización de las variables (diseño de medición)

Constructo	Tipo	Ítems	Escala	Contenido de los ítems
Ética Biocéntrica	Reflectivo	3	Likert de 5 puntos	Valor intrínseco de los seres vivos; consideración del ecosistema en las decisiones cotidianas; rechazo de la concepción instrumental de la naturaleza
Integración Curricular Ecopedagógica	Reflectivo	3		Integración transversal de contenidos ambientales; proyectos interdisciplinarios ecopedagógicos; vinculación de saberes locales y comunitarios
Conducta Proambiental Docente	Reflectivo	3		Actividades de cuidado ambiental en clase; promoción del reciclaje y la reducción de residuos; participación en campañas ambientales institucionales

Nota: Operacionalización de los constructos del modelo. Escalas elaboradas para este estudio a partir de: ética biocéntrica, Taylor (1986) y Stern et al. (1999); integración curricular ecopedagógica, Ruiz et al. (2021), Hung (2025) y Gallegos y Rhoades (2026); conducta proambiental docente, Lange (2024) y Sierra (2023). Las propiedades psicométricas (λ, α, ω, AVE) y la validez discriminante se reportan en las Tablas 2-3 y 2-4.

Las escalas se elaboraron para este estudio a partir de las definiciones conceptuales de sus marcos: la de ética biocéntrica, del principio de respeto por el valor inherente de los seres vivos (Taylor, 1986) y de la orientación biosférica (Stern et al., 1999); la de integración curricular, de la ecopedagogía (Ruiz et al., 2021; Hung, 2025), del enfoque ambiental transversal peruano (Gallegos y Rhoades, 2026) y de los referentes de

medición de la disposición profesional docente hacia la educación para el desarrollo sostenible (Vukelić, 2026); y la de conducta proambiental docente, de la investigación sobre qué se mide en este campo (Lange, 2024), de índices psicométricos latinoamericanos (Sierra, 2023) y de revisiones regionales de la conducta proambiental (Pontes et al., 2024).

La especificación del modelo siguió la lógica de estudios latinoamericanos que modelan factores cognitivos y actitudinales de la conducta proambiental (Jaime, 2026). Su validez de constructo y fiabilidad se evaluaron mediante análisis factorial confirmatorio en esta misma muestra.

2.3.4 Análisis

Tabla 2-2

Cálculo a priori del tamaño muestral por análisis

Análisis	Procedimiento	Parámetros	N mínimo (potencia ≥ 0,80)	Potencia con N = 375
Ajuste global del modelo CB-SEM	Potencia de la prueba de ajuste aproximado (MacCallum et al., 1996)	gl = 24; ε_0 = 0,05; ε_1 = 0,08; α = 0,05	375	0,801
Ecuación de la conducta proambiental (k = 2)	Prueba F, F no central (Cohen, 1988)	f^2 = 0,15; α = 0,05	68	> 0,999
Ecuación de la conducta proambiental (k = 2), cota conservadora	Prueba F, F no central (Cohen, 1988)	f^2 = 0,10; α = 0,05	100	> 0,999
Efecto indirecto a × b	Monte Carlo, significación conjunta, 2 000 réplicas	a = b = 0,26; c′ = 0,10; α = 0,05	160	> 0,99
Admisibilidad del modelo	N ≥ P + 1; gl = P(P + 1)/2 − q	P = 9; q = 21	10 (gl = 24 ≥ 0)	—

Nota: El tamaño de la muestra (N = 375) se fijó en el requerimiento del análisis más exigente (ajuste global). $\lambda = (N - 1) \cdot gl \cdot \varepsilon^2$ para la prueba de ajuste y $\lambda = f^2 \cdot N$ para la prueba F. Los valores de a, b y f^2 se sitúan por debajo del umbral de efecto moderado (Cohen, 1988) y de las asociaciones reportadas por Carlos et al. (2025) y Vásquez et al. (2022

Se estimó un modelo de ecuaciones estructurales basado en covarianzas por máxima verosimilitud con el paquete *semopy*.

1) Se evaluó el modelo de medición: cargas estandarizadas (λ), fiabilidad (α de Cronbach y ω de McDonald), varianza extraída media (AVE) y validez discriminante por el criterio de Fornell y Larcker (1981) y la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015).

2) Se estimó el modelo estructural con las rutas a (ética biocéntrica → integración curricular), b (integración curricular → conducta) y c′ (ética biocéntrica → conducta).

3) El efecto indirecto a × b y el efecto total se contrastaron con intervalos percentiles al 95 % por bootstrap de 5 000 réplicas (Hayes, 2022), y un efecto se consideró significativo si su intervalo excluía el cero.

4) El ajuste se informó mediante χ^2, χ^2/gl, CFI, TLI, RMSEA y SRMR, y se comparó con un modelo alternativo de un factor mediante la diferencia de χ^2. La estabilidad del efecto total se examinó con jackknife, submuestreo y bootstrap.

2.5 RESULTADOS

Los resultados se presentan en tres bloques: modelo de medición, modelo estructural y mediación, y estabilidad del efecto total.

2.5.1 Evaluación del modelo de medición

Las nueve cargas estandarizadas oscilaron entre 0,686 y 0,845, la fiabilidad compuesta entre ω = 0,796 y ω = 0,837, y la AVE entre 0,567 y 0,631 (Tabla 2-3). La raíz cuadrada de la AVE de cada constructo superó su mayor correlación con los demás y todos los valores HTMT se situaron por debajo de 0,85 (máximo = 0,674; Tabla 2-4).

Las medias de los tres constructos se ubicaron en el punto medio de la escala, con asimetría y curtosis cercanas a cero, y las correlaciones entre ellos fueron positivas y de magnitud moderada a alta (Tabla 2-5).

Tabla 2-3

Modelo de medición: propiedades psicométricas por ítem

Constructo	Ítem	λ	R^2	α	ω	AVE
Ética Biocéntrica	Considero que todos los seres vivos tienen valor intrínseco independientemente de su utilidad para el ser humano	.749	.561	.803	.800	.573
	Mis decisiones cotidianas toman en cuenta el impacto sobre el ecosistema	.686	.471	.803	.800	.573
	Rechazo la idea de que la naturaleza existe solo para satisfacer necesidades humanas	.830	.689	.803	.800	.573
Integración Curricular Ecopedagógica	Integro contenidos ambientales de forma transversal en mi planificación curricular	.795	.632	.834	.837	.631
	Diseño proyectos interdisciplinarios con enfoque ecopedagógico	.743	.552	.834	.837	.631
	Vinculo los saberes locales y comunitarios con la educación ambiental	.842	.709	.834	.837	.631
Conducta Proambiental Docente	Incorporo actividades de cuidado ambiental en mis sesiones de clase	.708	.501	.792	.796	.567
	Promuevo el reciclaje y la reducción de residuos entre mis estudiantes	.697	.486	.792	.796	.567
	Participo en campañas ambientales de mi institución educativa	.845	.714	.792	.796	.567

Nota: λ = carga factorial estandarizada; R^2 = correlación múltiple al cuadrado del ítem; α = alfa de Cronbach; ω = omega de McDonald; AVE = varianza extraída media, computadas en el análisis factorial confirmatorio. Criterios: λ ≥ 0,50; α y ω ≥ 0,70; AVE ≥ 0,50 (Fornell y Larcker, 1981).

Tabla 2-4

Validez discriminante (criterio de Fornell-Larcker y HTMT)

Constructo	1	2	3	√AVE	Máx. r	Fornell-Larcker
1. Ética Biocéntrica	**0.757**	0.563	0.674	0.757	0.641	Cumple
2. Integración Curricular Ecopedagógica	0.544	**0.794**	0.667	0.794	0.611	Cumple
3. Conducta Proambiental Docente	0.641	0.611	**0.753**	0.753	0.641	Cumple

Nota: Diagonal (negrita): raíz cuadrada de la AVE. Triángulo inferior: correlaciones entre constructos (Tabla 5). Triángulo superior: razón HTMT (Henseler et al., 2015); criterio HTMT < 0,85 (estricto) y < 0,90 (liberal); máximo HTMT = 0,674. Se cumple el criterio de Fornell y Larcker (1981) cuando la √AVE supera la mayor correlación del constructo.

Tabla 2-5

Estadísticos descriptivos y correlaciones

N.º	Variable	M	DE	Asimetría	Curtosis	1	2	3
1.	Ética Biocéntrica	3.001	0.664	-0.061	-0.040	—	0.641*** [0.577, 0.697]	0.544*** [0.468, 0.611]
2.	Conducta Proambiental Docente	3.000	0.667	-0.083	-0.404	0.641***	—	0.611*** [0.543, 0.670]
3.	Integración Curricular Ecopedagógica	3.000	0.667	-0.075	-0.157	0.544***	0.611***	—

Nota: N = 375. Triángulo superior: r con IC95 % por transformación Z de Fisher entre corchetes; triángulo inferior: r. Medias y desviaciones estándar en la escala Likert de 1 a 5. *p < .05; **p < .01; ***p < .001.

2.5.2 Modelo estructural y mediación

El modelo estructural reprodujo los datos con un $\chi^2(24) = 27{,}622$ ($p = 0{,}276$; $\chi^2/gl = 1{,}151$) y superó con amplitud al modelo alternativo de un factor ($\Delta\chi^2(3) = 306{,}638$; $p < .001$; Tabla 2-6).

Tabla 2-6

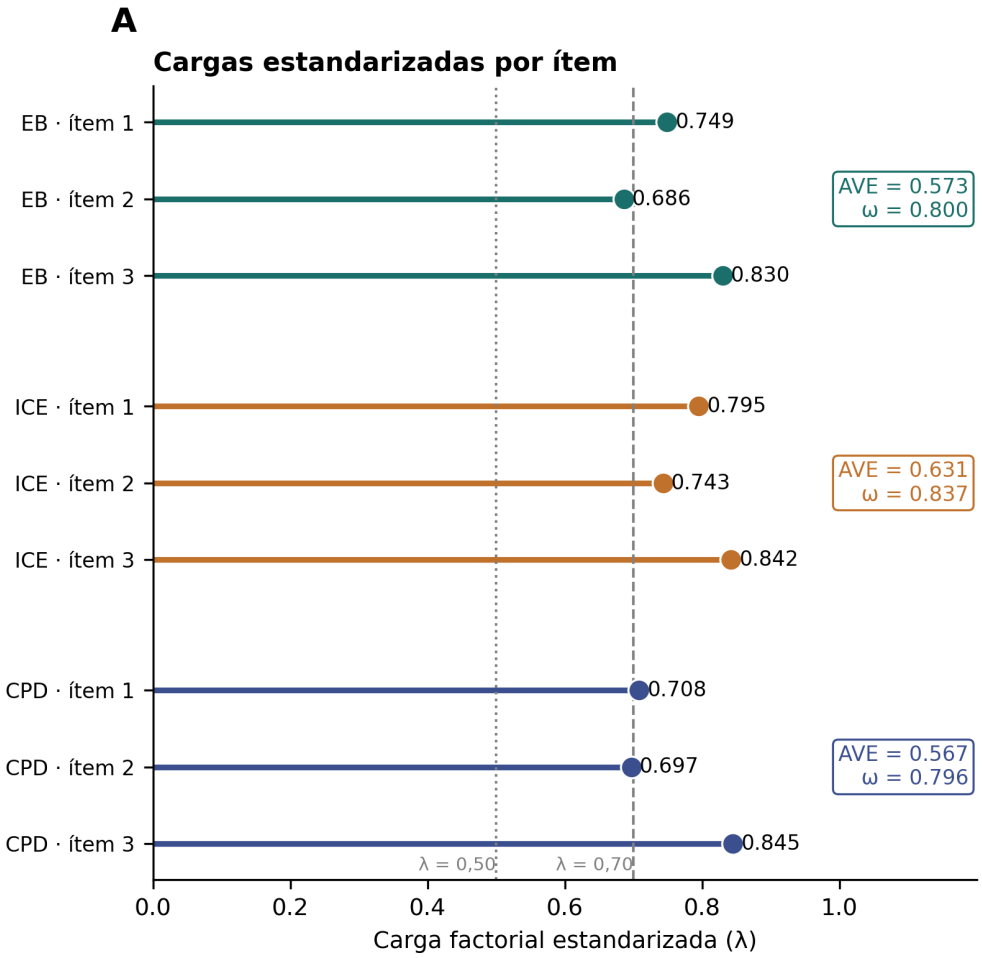
Ajuste global y comparación de modelos (CB-SEM)

Modelo	χ^2	gl	p	χ^2/gl	CFI	TLI	RMSEA	SRMR
Estructural hipotetizado (tres factores, mediación parcial)	27.622	24	0.276	1.151	0.997	0.996	0.020	0.0257
Alternativo (un factor)	334.260	27	< .001	12.380	0.786	0.714	0.174	0.119
Diferencia (alternativo – hipotetizado)	$\Delta\chi^2$ = 306.638	Δgl = 3	< .001	—	ΔCFI = −0.211	ΔTLI = −0.282	ΔRMSEA = +0.154	ΔSRMR = +0.093

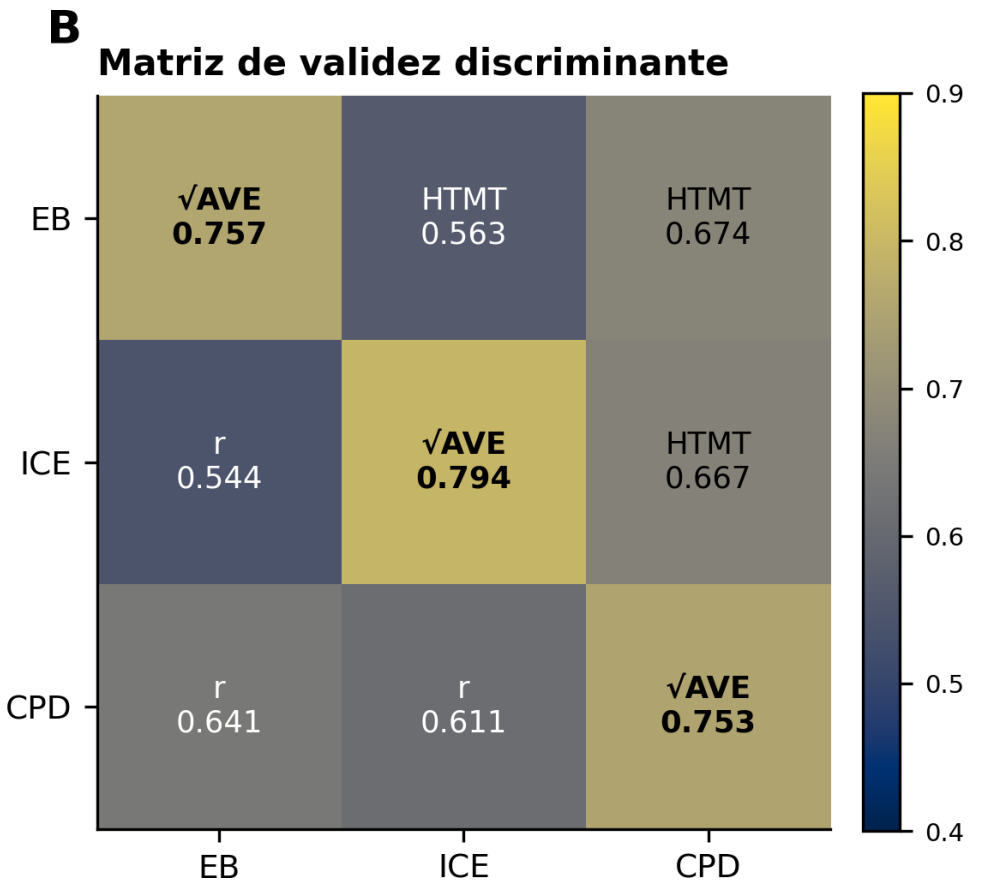
Nota: Estimación por máxima verosimilitud (N = 375). La diferencia de χ^2 contrasta los dos modelos anidados. Puntos de corte de referencia: CFI/TLI > 0,95, RMSEA < 0,06 y SRMR < 0,08.

Figura 2-1

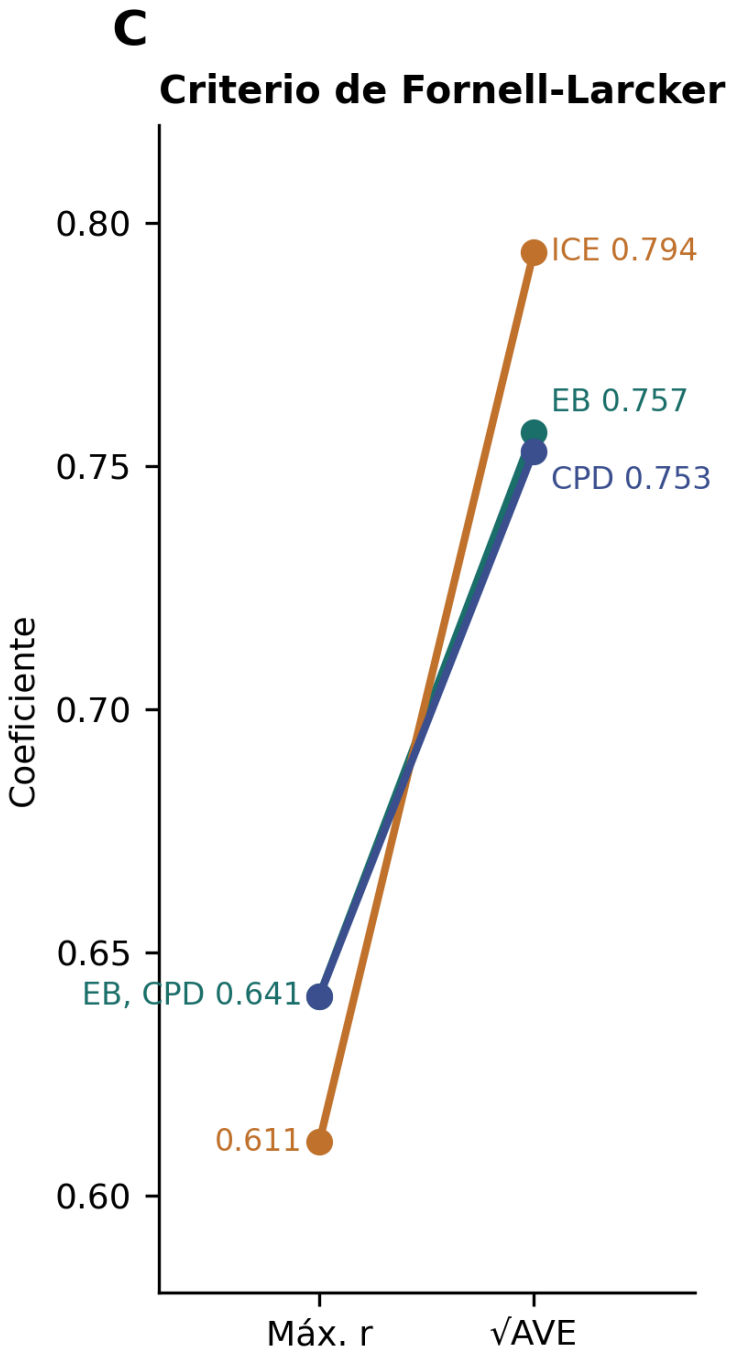
Modelo de medición: cargas estandarizadas, matriz de validez discriminante y criterio de Fornell-Larcker



Nota. Análisis factorial confirmatorio de tres factores (N = 375). Panel A: carga estandarizada (λ) de cada uno de los nueve ítems, agrupados por constructo (EB = ética biocéntrica; ICE = integración curricular ecopedagógica; CPD = conducta proambiental docente); rango λ = 0,686–0,845; líneas de referencia en λ = 0,50 y λ = 0,70; los recuadros indican la AVE y el ω de cada constructo (EB: AVE = 0,573, ω = 0,800; ICE: AVE = 0,631, ω = 0,837; CPD: AVE = 0,567, ω = 0,796).



Nota. Análisis factorial confirmatorio de tres factores (N = 375). Panel B: matriz mixta con la √AVE en la diagonal (0,757; 0,794; 0,753), las correlaciones entre constructos en el triángulo inferior (0,544; 0,641; 0,611) y la razón HTMT en el triángulo superior (0,563; 0,674; 0,667; máximo 0,674 < 0,85).



Nota. Análisis factorial confirmatorio de tres factores (N = 375). Panel C: comparación de la mayor correlación de cada constructo con su √AVE; la √AVE la supera en los tres constructos (EB y CPD: 0,641 frente a 0,757 y 0,753; ICE: 0,611 frente a 0,794), lo que cumple el criterio de Fornell-Larcker.

Las tres rutas directas fueron positivas y significativas, al igual que el efecto indirecto (Tabla 2-7).

Tabla 2-7

Rutas estructurales del modelo CB-SEM

	Ruta	β estandarizado	EE	p	Hipótesis
a	Ética Biocéntrica → Integración Curricular Ecopedagógica	0.518	0.070	< .001	H1
b	Integración Curricular Ecopedagógica → Conducta Proambiental Docente	0.397	0.065	< .001	H2
c′	Ética Biocéntrica → Conducta Proambiental Docente	0.420	0.068	< .001	H3
a × b	Ética Biocéntrica → Integración Curricular Ecopedagógica → Conducta Proambiental Docente (indirecto)	0.206	—	< .001	H4

Nota: β estandarizados del modelo estructural latente (máxima verosimilitud). EE = error estándar de la estimación no estandarizada, obteniendo p; los intervalos de las rutas a, b y c′, del efecto indirecto y del efecto total se presentan en la Tabla 8 (IC95 % bootstrap). *p < .05; **p < .01; ***p < .001.

La descomposición por *bootstrap* mostró que los intervalos de todos los efectos excluyeron el cero; el efecto indirecto explicó el 32,9 % del efecto total (0,2057/0,6255) y la vía directa conservó el 67,1 % (Tabla 2-8).

Tabla 2-8

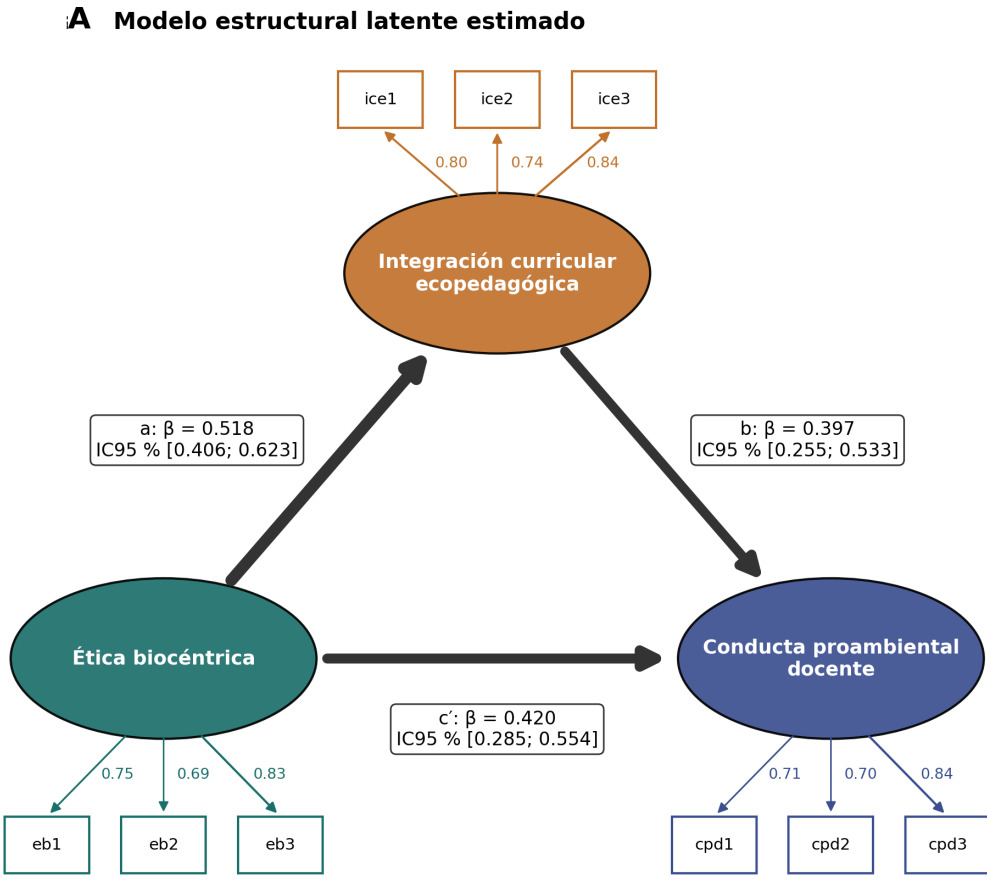
Descomposición de efectos (mediación, bootstrap 5000 réplicas)

	Efecto	Coeficiente	IC 95 %		Significativo
			inferior	superior	
a	(Ética biocéntrica → Integración curricular ecopedagógica)	0.518	0.406	0.623	Sí
b	(Integración curricular ecopedagógica → Conducta proambiental docente)	0.397	0.255	0.533	Sí
c′ directo	(Ética biocéntrica → Conducta proambiental docente)	0.420	0.285	0.554	Sí
a × b	(indirecto)	0.206	0.131	0.286	Sí
c total	(c′ + a × b)	0.625	0.527	0.720	Sí

Nota: Intervalos percentiles al 95 % obtenidos por bootstrap con 5 000 réplicas sobre el modelo latente (5 000 réplicas convergentes); significativo si el intervalo excluye el cero. Mediación parcial: c′ y a × b son significativos; proporción mediada = 0,2057/ 0,6255 = 32,9 %.

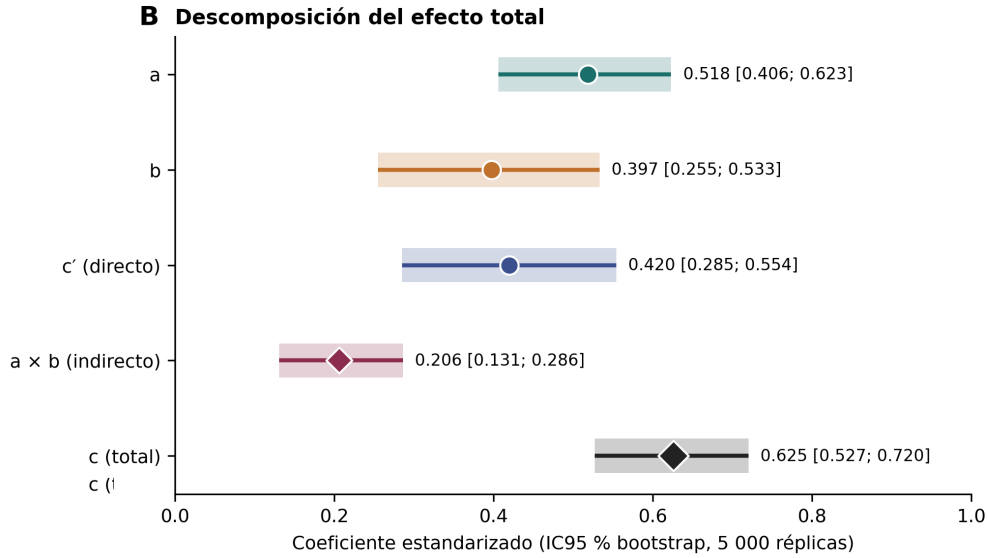
Figura 2-2

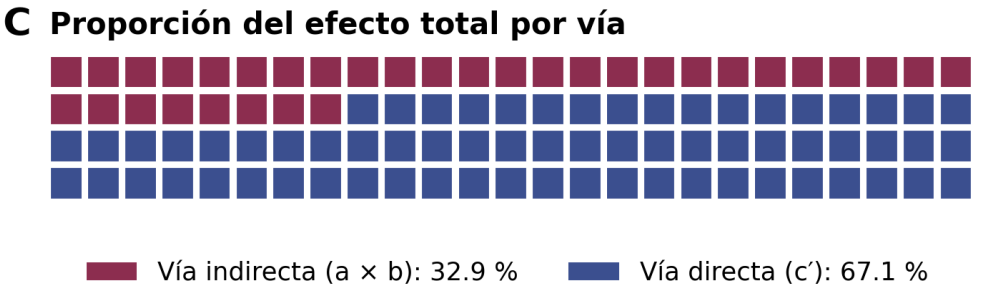
Modelo de medición: cargas estandarizadas, matriz de validez discriminante y criterio de Fornell-Larcker.



$\chi^2(24) = 27.622$; CFI = 0.997; TLI = 0.996; RMSEA = 0.020; SRMR = 0.0257; N = 375

Nota. Modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375); $\chi^2(24)$ = 27,622, CFI = 0,997, TLI = 0,996, RMSEA = 0,020, SRMR = 0,0257. Panel A: diagrama del modelo con las cargas estandarizadas de cada indicador y las rutas a (β = 0,518; IC95 % [0,406; 0,623]), b (β = 0,397; IC95 % [0,255; 0,533]) y c′ (β = 0,420; IC95 % [0,285; 0,554]); el grosor de cada flecha es proporcional a β.





Nota. Modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375); $\chi^2(24) = 27{,}622$, CFI = 0,997, TLI = 0,996, RMSEA = 0,020, SRMR = 0,0257. Panel B: estimaciones puntuales e intervalos de confianza al 95 % por bootstrap percentil con 5 000 réplicas de las rutas a, b y c′, del efecto indirecto a × b = 0,206 [0,131; 0,286] y del efecto total c = 0,625 [0,527; 0,720]; todos los intervalos excluyen el cero. Panel C: reparto del efecto total entre la vía indirecta (0,2057/0,6255 = 32,9 %) y la vía directa (67,1 %); cada cuadro representa un 1 % del efecto total.

Respecto de las hipótesis, H1 (β = +0,518; p < .001), H2 (β = +0,397; p < .001) y H3 (β = +0,420; p < .001) resultaron soportadas, así como H4, cuyo efecto indirecto (a × b = +0,206; IC95 % [0,131; 0,286]) excluyó el cero.

2.5.3 Estabilidad del efecto total

La asociación total entre ética biocéntrica y conducta proambiental docente conservó el signo positivo y la significación estadística en el 100 % de las re-muestras de los cuatro esquemas de perturbación, con coeficientes medios entre 0,578 y 0,586 (Tabla 2-9). Estos coeficientes, estimados por mínimos cuadrados sobre puntuaciones compuestas, fueron menores que el efecto total latente (0,625), dado que no corrigen el error de medida.

Tabla 2-9

Estabilidad del efecto total (jackknife, submuestreo y bootstrap)

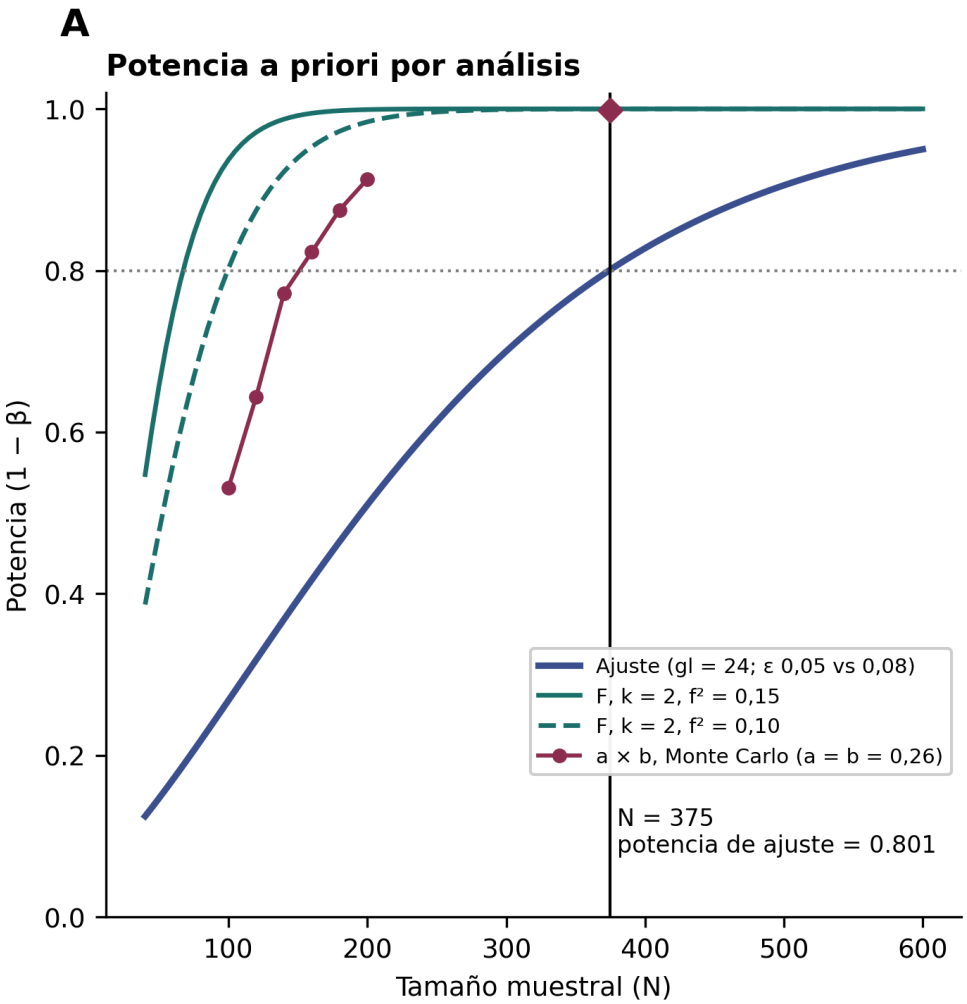
Prueba	Re-muestras	β medio	DE	iIC 95 %	Signo estable (%)	Significativo (%)
Jackknife (dejando uno fuera)	375	0.580	0.002	[0.5764, 0.5831]	100 %	100 %
Submuestreo 80 %	100	0.578	0.016	[0.5463, 0.6060]	100 %	100 %
Submuestreo 60 %	100	0.586	0.026	[0.5218, 0.6383]	100 %	100 %
Bootstrap (remuestreo con reemplazo)	1000	0.581	0.032	[0.5158, 0.6406]	100 %	100 %

Nota: Coeficiente estandarizado de la asociación total ética biocéntrica → conducta proambiental docente estimado por mínimos cuadrados ordinarios sobre puntuaciones compuestas como lente de robustez (β en la muestra completa = 0,580; p < .001). Porcentajes sobre el total de re-muestras de cada prueba..

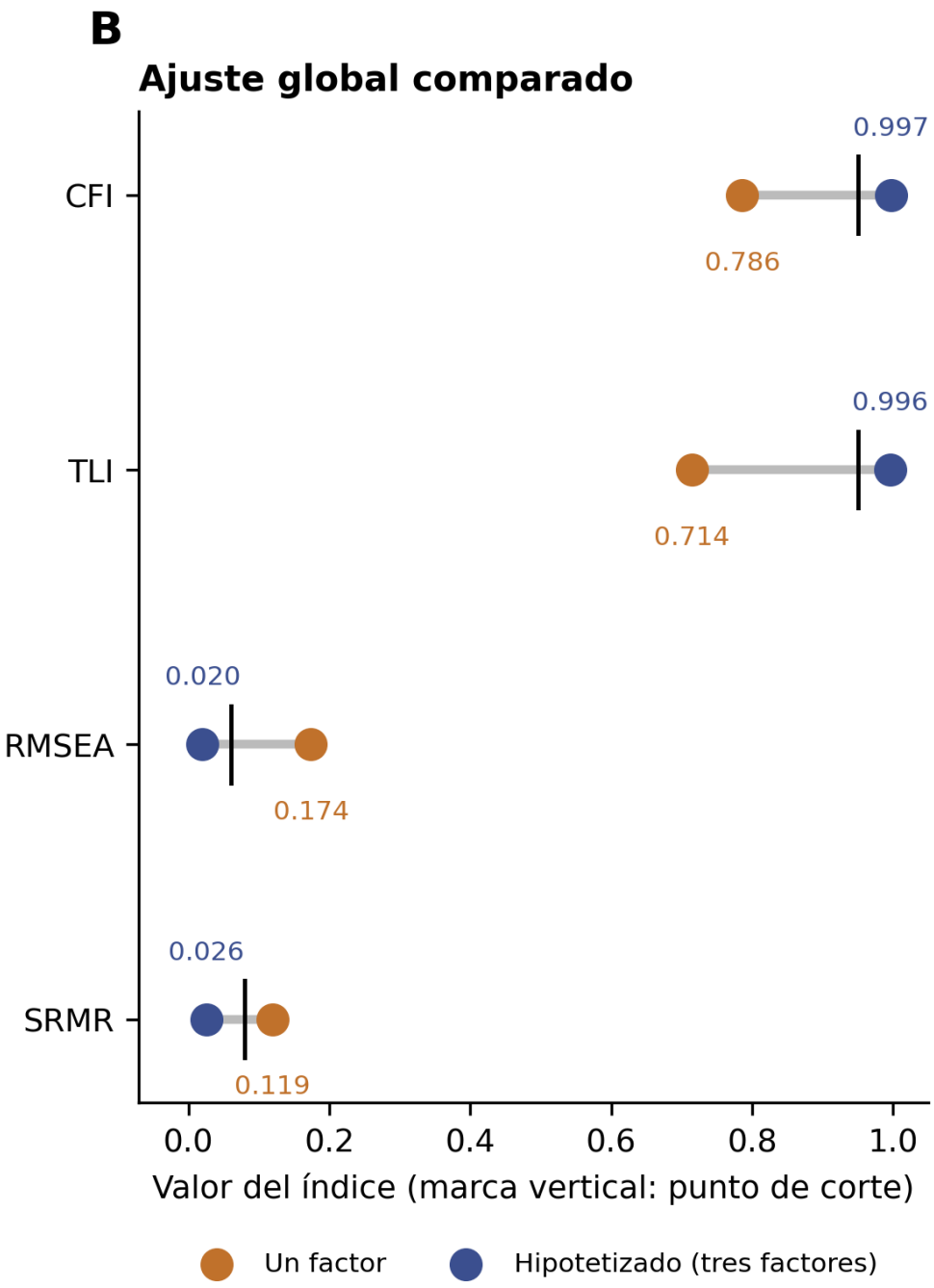
Como diagnósticos adicionales, el modelo estructural presentó un SRMR de 0,026 y el modelo de medición un SRMR de 0,036, ambos por debajo de 0,08.

Figura 2-3

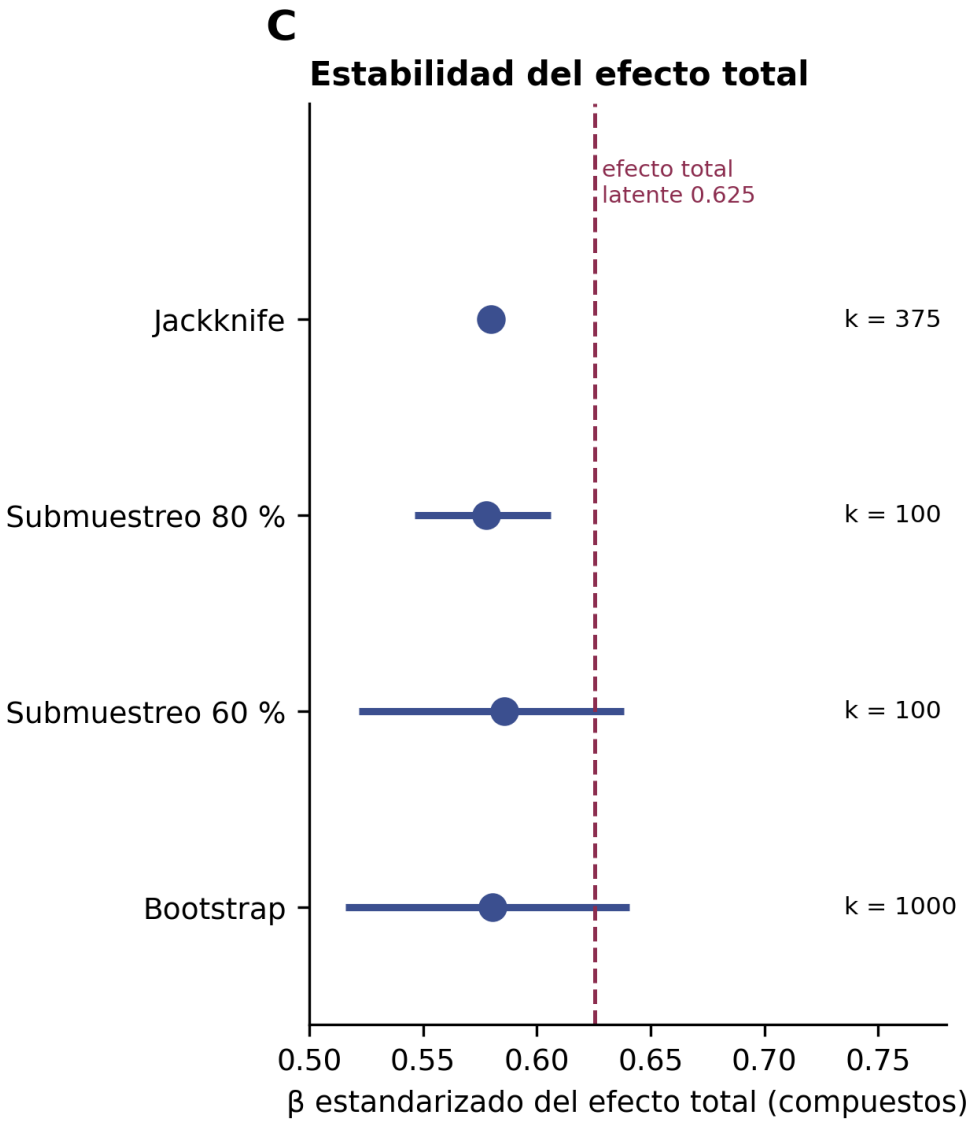
Potencia a priori, ajuste global comparado y estabilidad del efecto total



Nota. Cálculo de potencia a priori y modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375). Panel A: curvas de potencia de los análisis del estudio en función del tamaño muestral: prueba de ajuste aproximado del modelo (gl = 24; $\varepsilon_0 = 0{,}05$ frente a $\varepsilon_1 = 0{,}08$; $\alpha = 0{,}05$), prueba F de la ecuación de la conducta con dos predictores ($f^2 = 0{,}15$ y $f^2 = 0{,}10$) y potencia Monte Carlo del efecto indirecto (a = b = 0,26; 2 000 réplicas; significación conjunta); la línea vertical marca N = 375, donde la potencia de la prueba de ajuste es 0,801 y la del efecto indirecto supera 0,99; la línea punteada marca la potencia de 0,80.



Nota. Cálculo de potencia a priori y modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375). Panel B: índices de ajuste del modelo hipotetizado de tres factores (CFI = 0,997; TLI = 0,996; RMSEA = 0,020; SRMR = 0,026) frente al modelo alternativo de un factor (CFI = 0,786; TLI = 0,714; RMSEA = 0,174; SRMR = 0,119); las marcas verticales señalan los puntos de corte de referencia (CFI y TLI = 0,95; RMSEA = 0,06; SRMR = 0,08).



Nota. Cálculo de potencia a priori y modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375). Panel C: coeficiente estandarizado medio e IC95 % de la asociación total entre ética biocéntrica y conducta proambiental docente en cuatro esquemas de re-muestreo (jackknife, k = 375: 0,580; submuestreo del 80 %, k = 100: 0,578; submuestreo del 60 %, k = 100: 0,586; bootstrap, k = 1 000: 0,581), estimado por mínimos cuadrados sobre puntuaciones compuestas; la línea discontinua marca el efecto total latente (0,625), mayor porque corrige el error de medida.

2.5 DISCUSIÓN

El hallazgo más desconcertante no es que la integración curricular medie la relación entre ética y conducta, sino cuán poco de ese vínculo transita por ella. Si la ecopedagogía concibe el currículo como el espacio privilegiado donde los valores del docente se convierten en experiencia formativa, cabría esperar que la vía pedagógica concentrara la mayor parte del efecto; en cambio, solo un tercio del efecto total discurrió por la planificación y dos tercios se mantuvieron en una vía directa. Esta partición sugiere que la ética biocéntrica del profesorado opera, en primer lugar, como norma personal que orienta la acción también fuera del aula, de manera consistente con el principio de la teoría valor-creencia-norma según el cual las orientaciones biosféricas predisponen a la conducta (Stern et al., 1999) y con la evidencia de que las disposiciones docentes se expresan en acción mediante procesos próximos a ella, como la identidad ambiental (Felix y Clayton, 2025). La interpretación es necesariamente tentativa, porque la identidad ambiental no se midió y podría ocupar buena parte de lo que el modelo registra como vía directa.

La ruta de mayor magnitud fue, con todo, la que une la ética con la planificación curricular, y su lectura es menos paradójica. Que el compromiso con el valor intrínseco de la vida se asocie con una mayor integración de contenidos, proyectos y saberes locales es coherente con la propuesta de que el currículo sea un espacio de reconstrucción ética de la relación sociedad-naturaleza (Boxley, 2024; Hung, 2025), y con la noción de que los docentes que se perciben como promotores de una educación ambiental transformadora traducen esa convicción en prácticas de aula (Guevara et al., 2025). Desde esta perspectiva, la ética parece definir lo que el docente considera digno de enseñarse, y es en la decisión de qué enseñar donde el valor se vuelve pedagógicamente productivo.

El contraste más relevante con la literatura se refiere a la fuerza moderada de la ruta que conecta la integración curricular con la conducta, y a las razones por las que esta vía pedagógica no absorbe una proporción mayor del efecto. Una explicación plausible, aunque no contrastada aquí, radica

en que la planificación no se traduce en conducta en un vacío institucional. Los estudios sobre conducta proambiental en el contexto escolar identifican barreras y facilitadores que condicionan el paso de la preocupación a la acción (Huoponen, 2023), y la forma en que los centros conciben la sostenibilidad se relaciona con el comportamiento proambiental de la comunidad educativa (Cortés et al., 2026). En la misma línea, la evidencia latinoamericana sobre la certificación ambiental de establecimientos sugiere que las condiciones institucionales pueden modificar la conducta proambiental en el escenario escolar (Vásquez y Jiménez, 2026), y las perspectivas de docentes del sur de Chile sobre la educación ambiental al aire libre muestran que las prácticas más ambiciosas enfrentan desafíos logísticos y organizativos que ninguna convicción personal basta para resolver (Christie et al., 2025). Si estas condiciones filtran la ejecución de lo planificado, parte del efecto que en este estudio aparece como directo podría reflejar acciones que el docente emprende al margen de la estructura curricular, precisamente porque esta no siempre ofrece los medios para sostenerlas. Esta lectura es coherente con la identificación de motivos y predictores de la adopción de conductas proambientales por el profesorado (Öztürk y Pizmony, 2024), que apunta a una pluralidad de determinantes, y no solo curriculares, de la acción docente. El modelo, de un solo nivel, no puede discriminar esas condiciones, y sus resultados deben leerse como un patrón de asociaciones compatible con la mediación parcial, no como una descripción de procesos.

También conviene situar las magnitudes frente a la evidencia peruana. Las correlaciones entre los constructos (entre 0,544 y 0,641) son de un orden similar al reportado entre actitud proambiental y educación ambiental en docentes en formación de un instituto pedagógico público (Carlos et al., 2025), y la incidencia de la alfabetización ambiental sobre la conducta pedagógica sostenible documentada en docentes de Lima (Vásquez et al., 2022) ofrece una referencia convergente en dirección, aunque no en constructo ni en método. La comparación exige cautela: aquellos estudios trabajaron con otros indicadores y, en general, sin corregir el error de medida, mientras que aquí las rutas se estiman sobre variables latentes.

Este contraste metodológico tiene una consecuencia visible en la propia robustez: los coeficientes estimados sobre puntuaciones compuestas fueron sistemáticamente menores que el efecto total latente, lo que es consistente con la atenuación que produce ignorar el error de medida y sugiere que parte de la heterogeneidad entre estudios podría deberse al método de estimación y no solo a diferencias sustantivas entre poblaciones.

En términos de coincidencia, los resultados son congruentes con la evidencia de que los valores biosféricos y la identidad ambiental se asocian con la conducta proambiental (Wang et al., 2021) y con la aplicación de la teoría valor-creencia-norma a la adopción de estas conductas en la región (Hidalgo et al., 2023). La novedad radica en que esa asociación se descompone aquí en una porción normativa y otra pedagógica dentro del profesorado.

Las implicaciones apuntan en dos direcciones. Primero, las políticas de educación ambiental que actúan solo sobre las creencias del profesorado dejan sin activar una vía relevante, de modo que la competencia para traducir esas creencias en planificación constituye un objetivo formativo con identidad propia. Este argumento se alinea con las revisiones sobre el desarrollo profesional de docentes de ciencias en América Latina (Salazar y Peñaloza, 2023) y con la evidencia regional sobre las competencias docentes en educación para el desarrollo sostenible (Álvarez et al., 2024), que subrayan la necesidad de fortalecer la formación continua en esos ámbitos. Segundo, en sistemas con currículos de enfoque ambiental transversal, como el peruano (Gallegos y Rhoades, 2026), esa competencia es la que convierte un mandato normativo en experiencias de aprendizaje; si, como sugieren los datos, la vía directa pesa más que la pedagógica, la formación no puede limitarse a la sensibilización ética ni a la técnica de planificación, y debería articular ambas con las condiciones institucionales que permiten sostener la acción.

Conviene, por último, acotar el alcance de estas lecturas. La estabilidad del efecto total ante el jackknife, el submuestreo y el bootstrap indica que la estimación no depende de observaciones o submuestras particulares, pero

no la libera de las limitaciones de diseño: el carácter transversal impide fijar el orden temporal, el autorreporte de las tres medidas puede inflar las asociaciones y el muestreo por conveniencia restringe la generalización a los docentes analizados. Con esas reservas, el estudio aporta una estimación de la fracción del vínculo entre valores y conducta que transita por la práctica curricular, un parámetro hasta ahora postulado más que cuantificado.

2.6 CONCLUSIONES

La integración curricular ecopedagógica media parcialmente la relación entre la ética biocéntrica y la conducta proambiental de los docentes de educación básica: la vía pedagógica explica un tercio del efecto total y la vía directa, dos tercios. De ello se derivan tres implicaciones prácticas:

1) Incorporar en la formación inicial y continua del profesorado la planificación ecopedagógica contextualizada a la biodiversidad local como competencia explícita;

2) Acompañar las acciones de sensibilización ética con talleres de diseño de proyectos interdisciplinarios que vinculen saberes locales y comunitarios; y

3) Incluir indicadores de integración curricular ambiental en el seguimiento de la implementación del enfoque ambiental transversal.

El estudio presenta limitaciones que acotan sus conclusiones. El diseño transversal no permite establecer el orden temporal de las relaciones, por lo que el efecto indirecto debe interpretarse como un patrón de asociaciones compatible con la mediación y no como prueba causal. La selección por conveniencia impide generalizar las estimaciones al conjunto del profesorado. Las tres medidas son de autorreporte y proceden de una misma fuente, lo que puede inflar las asociaciones, y las escalas se elaboraron para este estudio, de modo que su estructura debe replicarse en muestras independientes. Por último, el análisis es de un solo nivel,

aunque los docentes pertenecen a instituciones distintas cuyo clima podría condicionar la integración curricular.

Quedan abiertas tres preguntas para investigaciones futuras: si la integración curricular mantiene su papel mediador en diseños longitudinales o de intervención; qué condiciones institucionales fortalecen la ruta entre ética e integración curricular; y si la formación inicial en ecopedagogía modifica el peso relativo de las vías directa e indirecta.

2.7 REFERENCIAS BIBLIOGRÁFICAS

Álvarez, A., Rieckmann, M., Lopera, M., & Aguirre, P. (2024). Teaching with A Rounder Sense of Purpose: a survey study on education for sustainable development competences in Latin America. *Frontiers in Education, 8,* 1205478. https://doi.org/10.3389/feduc.2023.1205478

Ashurov, A. (2026). Anthropocentric and Biocentric Theories in Environmental Ethics. *Metafizika Philosophy*, *1*(1), 79-90. https://doi.org/10.33864/mp.2026.v1.i1.79-90

Barzola, P., & Agoglia, O. (2024). Deep ecology and the philosophy of Emmanuel Levinas: the importance of moving from biocentric responsibility to environmental justice. *Ethics in Science and Environmental Politics, 24,* 31-45. https://doi.org/10.3354/esep00211

Boxley, S. (2024). Freire and Environmentalism: Ecopedagogy. *Journal of Philosophy of Education, 58*(6), 1051-1055. https://doi.org/10.1093/jopedu/qhae035

Carlos, O., Chachi, L., Salcedo, M., Janampa, R., Unocc, S., & Pacovilca, O. (2025). Actitud Proambiental y su relación con la Educación Ambiental en Estudiantes de un Instituto de Educación Superior

Pedagógico Público del Perú. *GnosisWisdom, 5*(3), 24-32. https://doi.org/10.54556/gnosiswisdom.v5i3.102

Christie, C., Cárcamo, L., Barraza, L., & Scheihing, E. (2025). Outdoor environmental education in Southern Chile: teachers' perspectives and challenges in excursions. *Environmental Education Research, 31*(5), 1069-1082. https://doi.org/10.1080/13504622.2025.2469760

Chávez, W. (2023). Educación ambiental y sostenibilidad en los Institutos Superiores Tecnológicos Públicos de la región Amazonas, Perú. *Revista Científica Pakamuros, 11*(1), 110-122. https://doi.org/10.37787/pakamuros-unj.v11i1.363

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.ª ed.). Lawrence Erlbaum.

Cortés, C., Alcántara, J., & Torres, J. (2026). Concepción de la sostenibilidad de los centros educativos y comportamiento proambiental de la comunidad educativa. *Revista de Investigación Educativa*(44). https://doi.org/10.6018/rie.627491

Felix, S., & Clayton, S. (2025). Describing teachers' environmental identity as part of education for sustainable development. *Environmental Education Research, 31*(7), 1389-1406. https://doi.org/10.1080/13504622.2025.2482740

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Gallegos, X., & Rhoades, J. (2026). Comparative assessment of the environmental education policies of Brazil, Colombia, Peru, and Venezuela. *Environmental Education Research, 32*(4), 1100-1127. https://doi.org/10.1080/13504622.2026.2616701

Gleiser, M. (2025). The Dawn of a Mindful Universe: A Pathway towards a Biocentric Global Ethics. *Zygon: Journal of Religion and Science*(0). https://doi.org/10.16995/zygon.18699

Guevara, I., Pérez, J., & Bravo, B. (2025). Teachers as promoters of transformative environmental education: perceptions and classroom practices. *Environmental Education Research, 31*(7), 1371-1388. https://doi.org/10.1080/13504622.2025.2478436

Hayes, A. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3.ª ed.). Guilford Press.

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hidalgo, J., Velastegui, A., Amaya, J., Soto, M., & Riel, A. (2023). The Role of Personality in the Adoption of Pro-Environmental Behaviors through the Lens of the Value-Belief-Norm Theory. *Sustainability, 15*(17), 12803. https://doi.org/10.3390/su151712803

Hung, R. (2025). Mapping ecopedagogy – or, How can we imagine an education for cobeings?. *Educational Philosophy and Theory*, 1-10. https://doi.org/10.1080/00131857.2025.2507902

Huoponen, A. (2023). From concern to behavior: barriers and enablers of adolescents' pro-environmental behavior in a school context. *Environmental Education Research, 30*(5), 677-699. https://doi.org/10.1080/13504622.2023.2180374

Jaime, J. (2026). Factores cognitivos y actitudinales asociados al comportamiento proambiental en estudiantes universitarios mexicanos. *Suma de Negocios, 17*(37), 145-153. https://doi.org/10.14349/sumneg/2026.v17.n37.a4

Lange, F. (2024). What is measured in pro-environmental behavior research?. *Journal of Environmental Psychology, 98*, 102381. https://doi.org/10.1016/j.jenvp.2024.102381

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

McIntosh, S., & Feltrin, A. (2025). Dismantling Human Supremacy: Ecopedagogy and Self-Rewilding as Pathways to Embodied Ethics and Cross-Species Solidarity. *Relations. Beyond Anthropocentrism, 12*(2). https://doi.org/10.7358/rela-2024-02-mcfe

McShane, K. (2024). A Basis for Biocentric Equality?. *Environmental Ethics, 46*(2), 131-148. https://doi.org/10.5840/enviroethics202451076

Öztürk, M., & Pizmony, O. (2024). Motives for and predictors of teachers' adoption of pro-environmental behaviours. Environmental Education Research, 31(3), 627-647. https://doi.org/10.1080/13504622.2024.2377367

Pedraja, L., Rodríguez, E., Muñoz, C., & Laroze, D. (2023). Sustainable Development Goals and Education: A Bibliometric Review—The Case of Latin America. *Sustainability*, *15*(12), 9833. https://doi.org/10.3390/su15129833

Pontes, L., Lauer, I., & Santos, J. (2024). Conexão com a natureza e comportamento pró-ambiental infantil: Revisão de pesquisas na América Latina. *Psicología desde el Caribe*, *40*(03), 99-121. https://doi.org/10.14482/psdc.40.3.720.589

Ruiz, S., Porcel, L., & Ruiz, A. (2021). La ecopedagogía en cuestión: una revisión bibliográfica. *Contextos Educativos. Revista de Educación*(28), 183-201. https://doi.org/10.18172/con.4489

Salazar, T., & Peñaloza, G. (2023). Professional development of science teachers in Latin America: a systematic review. *International Journal of Science Education*, *46*(10), 961-977. https://doi.org/10.1080/09500693.2023.2267747

Sierra, W. (2023). Propiedades psicométricas preliminares del Índice de Comportamiento Proambiental (ICP) en una muestra colombiana. *Informes Psicológicos*, *23*(1), 125-141. https://doi.org/10.18566/infpsic.v23n1a08

Stern, P., Dietz, T., Abel, T., Guagnano, G., & Kalof, L. (1999). A value-belief-norm theory of support for social movements: The case of environmentalism. *Human Ecology Review*, *6*(2), 81-97

Taylor, P. (1986). *Respect for nature: A theory of environmental ethics*. Princeton University Press

Tusa, F., Aguaded, I., & Tejedor, S. (2025). Values and Ethics as Education Policy: Media Framing of Ecuador's 2024 Curriculum Reform. *Education Sciences, 15*(10), 1328. https://doi.org/10.3390/educsci15101328

UNESCO. (2021). *Learn for our planet: A global review of how environmental issues are integrated in education.* UNESCO

Vukelić, N. (2026). Professional Readiness for Education for Sustainable Development: Development and Validation of the Teachers' Intention to Implement ESD Scale. *Education Sciences, 16*(6), 900. https://doi.org/10.3390/educsci16060900

Vásquez, M., & Jiménez, P. (2026). Efecto de la certificación ambiental de establecimientos educacionales en el conocimiento y conducta proambiental estudiantil. *Revista Colombiana de Educación*(101), e21554. https://doi.org/10.17227/rce.num101-21554

Vásquez, M., Yangali, J., Baldeón, M., & Huaita, D. (2022). Environmental Literacy and Its Impact on Sustainable Pedagogical Behaviors of Basic Education Teachers, Lima-Peru. *WSEAS TRANSACTIONS ON ENVIRONMENT AND DEVELOPMENT, 18*, 856-864. https://doi.org/10.37394/232015.2022.18.80

Wang, X., Van der Werff, E., Bouman, T., Harder, M., & Steg, L. (2021). I Am vs. We Are: How Biospheric Values and Environmental Identity of Individuals and Groups Can Influence Pro-environmental Behaviour. *Frontiers in Psychology, 12*, 618956. https://doi.org/10.3389/fpsyg.2021.618956

CAPÍTULO III

GESTIÓN DE LA CALIDAD PEDAGÓGICA, MEDIACIÓN TECNOLÓGICA DOCENTE Y CALIDAD DEL APRENDIZAJE HÍBRIDO: UNA CONTINGENCIA TERRITORIAL EN EL SUR DEL PERÚ

Luis Enrique Rivas Pacherrez

Universidad Nacional de Piura

Ana Consuelo Yovera Espinoza

Universidad Nacional de Piura

Yonny Marvin Navarro Yanayaco

Universidad Nacional de Piura

Noelia Elizabeth Palacios Lopez

Universidad César Vallejo

Resumen

La consolidación de la educación híbrida exige comprender cómo la gestión institucional se convierte en aprendizajes de calidad cuando la enseñanza depende de la mediación tecnológica del docente, y si esa conversión varía entre territorios. Este estudio estimó la asociación entre la gestión de la calidad pedagógica y la calidad del aprendizaje híbrido, el papel mediador de la mediación tecnológica docente y la diferencia del efecto entre zonas urbanas y rurales. Con un diseño cuantitativo, no experimental y transversal, se encuestó en 2025 a 375 docentes de secundaria y universidad de Arequipa, Cusco y Puno (196 urbanos y 179 rurales), mediante muestreo por conveniencia. Se estimó un modelo de ecuaciones estructurales por máxima verosimilitud, con bootstrap de 5000 réplicas, invarianza de medición y un contraste multigrupo de pendientes con errores robustos. El modelo mostró buen ajuste y validez discriminante, y la invarianza entre zonas se sostuvo. La gestión se asoció directamente con la calidad del aprendizaje ($c' = .318$) y de forma indirecta mediante la mediación tecnológica ($a \times b = .210$; IC 95 % [.116, .322]), que transmitió el 39.8 % del efecto total. La pendiente fue mayor en la zona urbana ($\beta = .482$) que en la rural ($\beta = .358$; $\Delta\beta = .123$, $p = .043$), con medias equivalentes entre zonas. Los resultados sugieren que la brecha educativa digital no es solo de acceso, sino también de eficacia de la gestión pedagógica, con implicaciones para el aseguramiento de la calidad híbrida y la política educativa territorial.

Palabras clave: gestión de la calidad pedagógica; mediación tecnológica docente; aprendizaje híbrido; brecha territorial; ecuaciones estructurales; Perú.

Abstract

The consolidation of hybrid education requires understanding how institutional management is converted into quality learning when teaching depends on the technological mediation exercised by teachers, and whether this conversion varies

across territories. This study estimated the association between pedagogical quality management and hybrid learning quality, the mediating role of teachers' technological mediation, and the difference in this effect between urban and rural areas. Using a quantitative, non-experimental, cross-sectional design, 375 secondary and university teachers from Arequipa, Cusco and Puno (196 urban and 179 rural) were surveyed in 2025 through convenience sampling. A structural equation model was estimated by maximum likelihood, with a 5000-replicate bootstrap, measurement invariance testing and a multigroup slope contrast with robust errors. The model showed good fit and discriminant validity, and invariance across zones held. Management was directly associated with learning quality ($c' = .318$) and indirectly through technological mediation ($a \times b = .210$; 95% CI [.116, .322]), which transmitted 39.8% of the total effect. The slope was steeper in urban ($\beta = .482$) than in rural areas ($\beta = .358$; $\Delta\beta = .123$, $p = .043$), with equivalent means across zones. The results suggest that the digital education gap is not only one of access but also one of the effectiveness of pedagogical management, with implications for quality assurance in hybrid education and for territorial education policy.

Keywords: *pedagogical quality management; teachers' technological mediation; hybrid learning; territorial gap; structural equation modeling; Peru.*

3.1 INTRODUCCIÓN

La consolidación de la modalidad híbrida tras la emergencia sanitaria desplazó el problema de la calidad educativa desde la disponibilidad de plataformas hacia la capacidad de las instituciones para gobernar pedagógicamente entornos que combinan presencialidad y virtualidad (Clifft y Assiouras, 2023). La experiencia del aprendizaje digital remoto transformó la enseñanza híbrida, pero las perspectivas de los docentes en países en desarrollo revelan desafíos persistentes de implementación (Brenya, 2024), y la calidad del aprendizaje en estos entornos depende de que se fomente la autonomía del estudiante desde marcos pedagógicos explícitos (Adigun et al., 2025). Los sistemas de aseguramiento de la calidad diseñados para la enseñanza presencial resultan insuficientes: la

literatura sobre calidad en línea y combinada insiste en que los estándares, el monitoreo y la mejora continua deben reconfigurarse para la docencia mediada por tecnología (Greere y Crozier, 2024), y sitúa entre los indicadores clave la interacción docente-estudiante, la retroalimentación oportuna, el apoyo al estudiante y la aplicación planificada de la tecnología (Hafeez et al., 2022). La cuestión ya no es si la tecnología está presente, sino cómo la gestión de la calidad pedagógica se convierte en aprendizajes de calidad cuando la enseñanza depende de la mediación tecnológica que ejerce el docente. Por ello, la calidad del aprendizaje híbrido constituye hoy un resultado que debe explicarse desde la gestión institucional y no solo desde la tecnología disponible.

La investigación sobre liderazgo y gestión escolar ha mostrado que los efectos de la gestión institucional sobre la calidad de la enseñanza operan en buena medida a través de mecanismos docentes, pero esa tradición se ha desarrollado sobre todo en contextos presenciales. No ha examinado la mediación tecnológica docente —la selección de recursos digitales con propósito pedagógico, el andamiaje en plataformas y la adaptación del diseño instruccional a la conectividad disponible— como mecanismo por el cual la gestión de la calidad alcanza al aprendizaje híbrido. Existen indicios de que la tecnología rinde según cómo se gestiona pedagógicamente: la relación entre recursos tecnológicos, uso de las TIC y eficiencia de las escuelas de la región se ha analizado con los datos de PISA 2018 (Agasisti et al., 2023), línea que sugiere que el rendimiento de los recursos depende de su gestión, y en el Perú la competencia digital docente se asocia de forma positiva y significativa con las estrategias de aprendizaje en el aula (Pérez et al., 2026). Su articulación con la gestión institucional permanece, sin embargo, poco explorada.

Un segundo vacío concierne a las brechas territoriales. En América Latina las desigualdades de oportunidad educativa tienen una marcada expresión territorial (Borchers y Cunha, 2025), la brecha digital se cruza con la educación y con el género (Ancheta et al., 2021) y las políticas de conectividad adoptadas durante la COVID-19 condicionaron la infraestructura de la que depende la educación híbrida (Califano y Becerra,

2021). La educación virtual posterior a la emergencia enfrentó, además, desafíos de organización de la enseñanza y de evaluación, según una encuesta a 2742 estudiantes y 926 docentes de Chile, México y Ecuador (Remesal y Villarroel, 2023). El Perú es un sitio de prueba exigente. La expansión de la banda ancha ha sido desigual entre territorios (Bonifaz y Aguirre, 2026); los candidatos a plazas docentes permanentes prefieren escuelas cercanas y urbanas, lo que condiciona la dotación de las escuelas rurales (Bertoni et al., 2023); la estrategia de educación a distancia «Aprendo en casa» exigió implementar el enfoque intercultural en escuelas rurales mediante la mediación docente (De la Portilla, 2023); y la calidad percibida de la educación virtual en una universidad pública de Puno es heterogénea, pues el 37 % de los estudiantes percibe una calidad media y el 19 % la califica como insatisfactoria (Justo et al., 2025). Se desconoce si la conversión de la gestión de la calidad pedagógica en calidad del aprendizaje híbrido opera con la misma intensidad en zonas urbanas y rurales.

El objetivo de este estudio es estimar la asociación entre la gestión de la calidad pedagógica y la calidad del aprendizaje híbrido en docentes de educación secundaria y universitaria del sur peruano, mediante tres hipótesis: un efecto directo (H1), un efecto indirecto a través de la mediación tecnológica docente (H2) y una diferencia de ese efecto entre zonas urbanas y rurales (H3). Para ello se estima un modelo de ecuaciones estructurales con tres constructos latentes, se verifica la equivalencia de la medición entre zonas y se contrasta la diferencia de pendientes, de modo que una eventual brecha no pueda atribuirse a que los docentes interpreten los constructos de forma distinta. Más allá del caso peruano, el estudio aporta a la teoría de la gestión y el liderazgo pedagógico un mecanismo de conversión propio de la modalidad híbrida y una condición de contorno territorial que permite distinguir si la brecha educativa digital es una brecha de acceso o una brecha en la eficacia de la gestión, distinción trasladable a otros sistemas con desigualdades territoriales de conectividad. El artículo presenta a continuación el marco teórico y las hipótesis, el método, los resultados y su discusión.

3.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

La teoría de la gestión de la calidad educativa concibe la calidad pedagógica como el resultado de procesos institucionales de planificación con estándares explícitos, monitoreo sistemático de la enseñanza y uso de la información en ciclos de mejora (Greere y Crozier, 2024; Hafeez et al., 2022). En esta perspectiva, la gestión de la calidad pedagógica es un antecedente organizacional que estructura las condiciones en que se enseña y que, por tanto, debería asociarse con la calidad del aprendizaje en las sesiones híbridas, tal como se ha documentado para el liderazgo que moviliza la autonomía y la innovación docentes y para el liderazgo curricular (Hsieh et al., 2025; Kuloğlu et al., 2026).

H1. La gestión de la calidad pedagógica se relaciona de forma positiva y directa con la calidad del aprendizaje híbrido.

La perspectiva de la mediación tecnológica sostiene que los recursos digitales no actúan por sí mismos sobre el aprendizaje, sino a través de la acción docente que los selecciona, adapta e integra con propósito pedagógico, un repertorio descrito por los marcos de competencia digital y por la evidencia metaanalítica sobre la adopción docente de la tecnología (Scherer et al., 2019; Rubio et al., 2023). La gestión de la calidad orienta y retroalimenta esas decisiones; cabe esperar, por tanto, que una gestión más sólida se exprese en una mediación tecnológica más pertinente y que esta eleve la calidad del aprendizaje, en línea con los modelos de mediación documentados para otros mecanismos docentes como la eficacia colectiva, la autonomía y la innovatividad (Özdemir et al., 2025; Demir y Kalman, 2025; Long y Xia, 2025).

H2. La mediación tecnológica docente es una variable mediadora que transmite una parte del efecto de la gestión de la calidad pedagógica sobre la calidad del aprendizaje híbrido (efecto indirecto positivo).

La conversión de la gestión en aprendizaje depende, además, de las condiciones materiales de la docencia híbrida. Donde la conectividad y el soporte técnico son más precarios, los lineamientos de calidad

encuentran más obstáculos para materializarse en el aula: las políticas de conectividad han sido desiguales (Califano y Becerra, 2021), la expansión de la banda ancha en el Perú no ha sido homogénea (Bonifaz y Aguirre, 2026) y la capacidad de la gestión pedagógica para traducirse en calidad de enseñanza puede debilitarse en contextos rurales (Lin y Hu, 2026). La zona de la institución actúa así como variable de agrupación del contraste multigrupo.

H3. El efecto de la gestión de la calidad pedagógica sobre la calidad del aprendizaje híbrido es mayor en zonas urbanas que en zonas rurales.

3.3 MATERIALES Y MÉTODOS

3.3.1 Diseño

Se adoptó un diseño cuantitativo, no experimental, transversal y explicativo, semejante al de estudios predictivos transversales recientes sobre el aprendizaje en línea en universitarios peruanos (Concha et al., 2025), desarrollado durante el año académico 2025 en instituciones de educación secundaria y universitaria con modalidad híbrida de Arequipa, Cusco y Puno. La unidad de análisis fue el docente, con análisis en un solo nivel.

3.3.2 Participantes y tamaño de la muestra

El tamaño muestral se fijó a priori con el análisis de potencia para la prueba de ajuste aproximado de MacCallum et al. (1996). Con gl = 24 (tres constructos latentes con tres indicadores cada uno y las rutas a, b y c′), $\alpha = .05$, $\varepsilon 0 = .05$ y $\varepsilon 1 = .08$, N = 375 arroja una potencia de .801, con valor crítico de χ^2 de 67.14, mientras que N = 374 queda por debajo del umbral (.7998). Esa N cubre también las dos pruebas F de regresión ($\alpha = .05$; $1 - \beta = .80$): las dos rutas que convergen en la calidad del aprendizaje híbrido (gl = 2 y 372) exigen $f^2 \geq 0.026$ y la diferencia de pendientes entre zonas (término grupo × gestión; gl = 1 y 371) exige $f^2 \geq 0.021$. Ambos umbrales quedan por debajo de las correlaciones agregadas de la evidencia

metaanalítica sobre adopción tecnológica docente, entre r = .33 y r = .59 (k = 114 estudios; N = 34 357 docentes), equivalentes a f^2 entre 0.12 y 0.53 (Scherer et al., 2019). El muestreo fue no probabilístico por conveniencia, condicionado por el acceso institucional a docentes en ejercicio híbrido; por ello no se afirma representatividad poblacional. Se incluyó a docentes con al menos un periodo académico de enseñanza híbrida, y se excluyeron los cuestionarios incompletos. La muestra comprendió 196 docentes de zona urbana y 179 de zona rural.

3.3.3 Instrumentos

Los tres constructos se midieron con escalas Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo) y tres indicadores reflectivos cada una (Tabla 3-1). Los ítems se elaboraron para este estudio a partir de indicadores de calidad de la enseñanza en línea y combinada para la gestión, de los marcos de competencia digital docente para la mediación tecnológica (Quast et al., 2025; Rubio et al., 2023) y de los indicadores de calidad del aprendizaje combinado para el resultado (Batista y Gavilan, 2023; Panigrahi et al., 2024). La experiencia docente en modalidad híbrida (años) se registró como covariable y la zona de la institución como variable de agrupación. El cuestionario se administró en 2025.

Tabla 3-1

Operacionalización de las variables

Constructo	Tipo	Ítems	Escala	Indicadores
Gestión de la calidad pedagógica (GCP)	Reflectivo	3	Likert de 5 puntos	GCP1 planificación con estándares explícitos; GCP2 monitoreo directivo con retroalimentación; GCP3 uso del monitoreo en ciclos de mejora
Mediación tecnológica docente (MTD)				MTD1 selección de recursos según propósito y conectividad; MTD2 andamiaje sincrónico y asincrónico en plataformas; MTD3 adaptación del diseño instruccional a condiciones limitadas

Constructo	Tipo	Ítems	Escala	Indicadores
Calidad del aprendizaje híbrido (CAH)	Reflectivo	3	Likert de 5 puntos	CAH1 logro de los aprendizajes previstos; CAH2 participación sostenida de los estudiantes; CAH3 retroalimentación oportuna y pertinente
Zona de la institución	Agrupación (nominal)	—	Rural (n = 179); Urbana (n = 196)	Ubicación de la institución

Nota: Las propiedades psicométricas de los indicadores se presentan en la Tabla 2.

3.3.4 Análisis de datos

Se estimó un modelo de ecuaciones estructurales basado en covarianzas por máxima verosimilitud (semopy, Python):

1) Primero se evaluó el modelo de medición (cargas estandarizadas, α, ω y varianza extraída media, AVE) y la validez discriminante con la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015) y el criterio de Fornell-Larcker.

2) Segundo, el ajuste global con χ^2, CFI, TLI, RMSEA y SRMR (criterio < .08; Hu y Bentler, 1999), frente a un modelo alternativo de un factor que sirve además como prueba de un factor común de método, y la invarianza de medición entre zonas con modelos anidados configural, métrico y escalar, donde una disminución del CFI ≥ .010 indicaría falta de invarianza (Cheung y Rensvold, 2002).

3) Tercero, las rutas estructurales y el efecto indirecto a × b con intervalos de confianza al 95 % por bootstrap de 5000 réplicas (mediación parcial si c′ permanece significativo). Cuarto, tras comparar medias (prueba t de Welch) y varianzas (prueba de Levene) entre zonas, H3 se contrastó mediante la diferencia de pendientes de la gestión sobre la calidad

del aprendizaje (Δβ), probada con el término grupo × gestión en una regresión por mínimos cuadrados sobre las puntuaciones de los constructos, con errores robustos HC3.

4) Por último, la estabilidad del efecto total de la gestión se examinó con jackknife, submuestreo al 80 % y al 60 % y remuestreo Monte Carlo, y la robustez de las rutas con una regresión jerárquica que incorpora la experiencia docente híbrida.

3.4 RESULTADOS

3.4.1 Modelo de medición

Las cargas estandarizadas oscilaron entre .687 y .831, y la fiabilidad y la validez convergente cumplieron los criterios de referencia (Tabla 3-2).

Tabla 3-2

Modelo de medición: propiedades psicométricas por ítem

Constructo	Ítem	λ	R^2	α	ω	AVE
Gestión de la calidad pedagógica	GCP1	.779	.607	.840	.839	.635
	GCP2	.798	.637	.840	.839	.635
	GCP3	.814	.663	.840	.839	.635
Calidad del aprendizaje híbrido	CAH1	.687	.472	.787	.788	.555
	CAH2	.709	.503	.787	.788	.555
	CAH3	.831	.691	.787	.788	.555
Mediación tecnológica docente	MTD1	.742	.551	.793	.793	.560
	MTD2	.732	.536	.793	.793	.560
	MTD3	.771	.594	.793	.793	.560

Nota: Códigos de ítem según la Tabla 1. λ = carga estandarizada del análisis factorial confirmatorio; R^2 = correlación múltiple al cuadrado; AVE = varianza extraída media. Referencias: λ ≥ .50; α y ω ≥ .70; AVE ≥ .50.

Ambos criterios de validez discriminante se cumplieron: el HTMT máximo fue .647 (< .85) y la raíz cuadrada de la AVE de cada constructo superó su mayor correlación con los demás (Tabla 3-3).

Tabla 3-3

Validez discriminante: HTMT y criterio de Fornell-Larcker

Panel	Par o constructo	Valor	Referencia	**Cumple**
A. HTMT	GCP — CAH	.578	< .85	Sí
	GCP — MTD	.647	< .85	Sí
	CAH — MTD	.586	< .85	Sí
B. Fornell-Larcker	GCP: √AVE	.797	r máx. = .641	Sí
	CAH: √AVE	.745	r máx. = .594	Sí
	MTD: √AVE	.748	r máx. = .641	Sí

Nota: Códigos según la Tabla 3-1. Panel A: razón heterorrasgo-monorrasgo (Henseler et al., 2015). Panel B: raíz cuadrada de la AVE frente a la mayor correlación del constructo con los demás (Tabla 3-4).

3.4.2 Estadísticos descriptivos y ajuste del modelo

Las escalas mostraron medias próximas al punto medio y asimetría entre –0.099 y 0.033. Las correlaciones entre constructos fueron positivas y significativas (r entre .570 y .641), y la experiencia docente híbrida no se correlacionó significativamente con ninguno de ellos (Tabla 3-4).

Tabla 3-4

Estadísticos descriptivos y correlaciones

N.º	Variable	M	DE	Asimetría	Curtosis	1	2	3	4
1.	Gestión de la calidad pedagógica	3.000	0.665	–0.029	–0.284	—	.570*** [.498, .635]	.641*** [.578, .697]	.101 [–.000, .200]
2.	Calidad del aprendizaje híbrido	3.000	0.667	–0.099	–0.459	.570***	—	.594*** [.524, .656]	.099 [–.002, .198]
3.	Mediación tecnológica docente	3.001	0.664	0.033	–0.075	.641***	.594***	—	.062 [–.039, .163]
4.	Experiencia docente híbrida (años)	5.905	4.775	1.492	1.662	.101	.099	.062	—

Nota: N = 375. Triángulo superior: r con IC 95 % (Z de Fisher); triángulo inferior: r. Medias de los constructos en la escala de 1 a 5. *** p < .001.

El modelo hipotetizado reprodujo la estructura de covarianzas con $\chi^2(24) = 14.03$, $p = .946$, y un SRMR inferior a .08. Como el χ^2 resultó inferior a sus grados de libertad, la no centralidad estimada es nula, de modo que el RMSEA toma el valor 0 y el CFI el valor 1 por definición. El modelo alternativo de un factor mostró un ajuste claramente inferior. En la invarianza de medición entre zonas, del modelo configural al métrico y al escalar el CFI no disminuyó y el RMSEA no aumentó (Tabla 3-5).

Tabla 3-5

Ajuste global del modelo e invarianza de medición entre zonas

Modelo	χ^2	gl	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA
Estructural hipotetizado	14.03	24	1.000	1.000	.000	.017	—	—
Alternativo de un factor	328.06	27	.778	.704	.173	.122	—	—
Invarianza configural	—	—	.988	.981	.031	—	—	—
Invarianza métrica	—	—	.999	.999	.010	—	+.012	−.021
Invarianza escalar	—	—	1.000	1.000	.007	—	.000	−.003

Nota: CFI, TLI y RMSEA se informan sin punto de corte; criterio para el SRMR: < .08 (Hu y Bentler, 1999). En invarianza, Δ respecto del nivel anterior; una disminución del CFI ≥ .010 indicaría falta de invarianza (Cheung y Rensvold, 2002).

3.4.3 Modelo estructural y mediación

Tabla 3-6

Rutas estructurales del modelo CB-SEM

Ruta	β	EE	p	Hipótesis
GCP → MTD (a)	.607	.068	< .001	—
MTD → CAH (b)	.346	.080	< .001	—
GCP → CAH (c′)	.318	.074	< .001	H1 soportada
GCP → MTD → CAH (indirecto a × b)	.210	—	< .001	H2 soportada

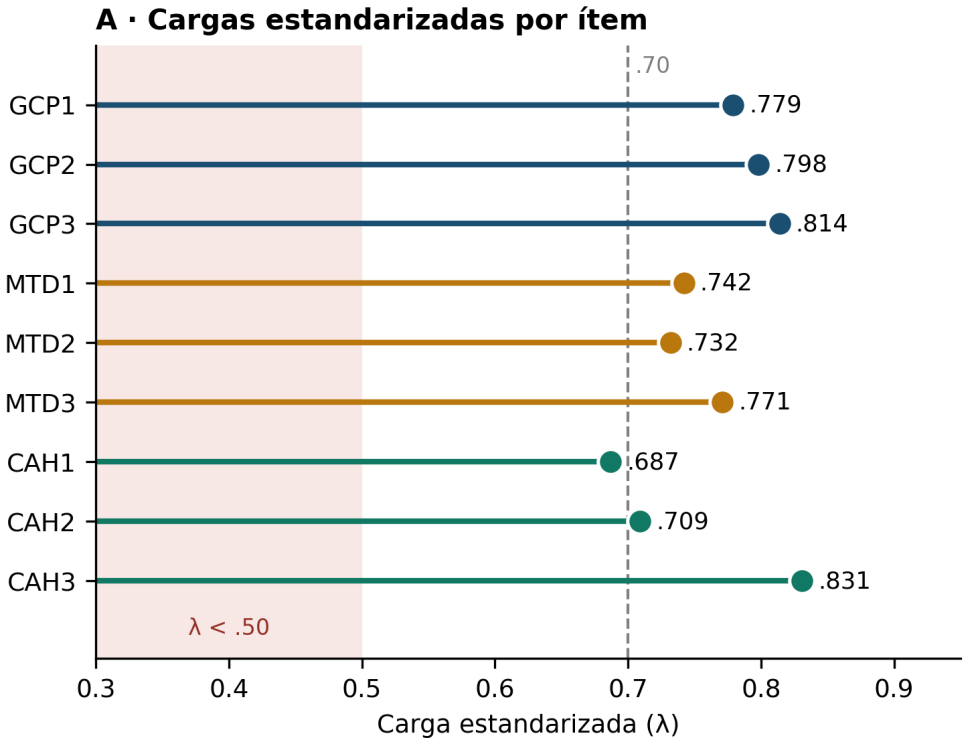
Nota: Códigos según la Tabla 3-1. β estandarizados del modelo estructural latente (máxima verosimilitud); EE = error estándar. El intervalo de confianza del efecto indirecto figura en la Tabla 3-7

La gestión de la calidad pedagógica se asoció positivamente con la mediación tecnológica docente (ruta a), y esta con la calidad del aprendizaje híbrido

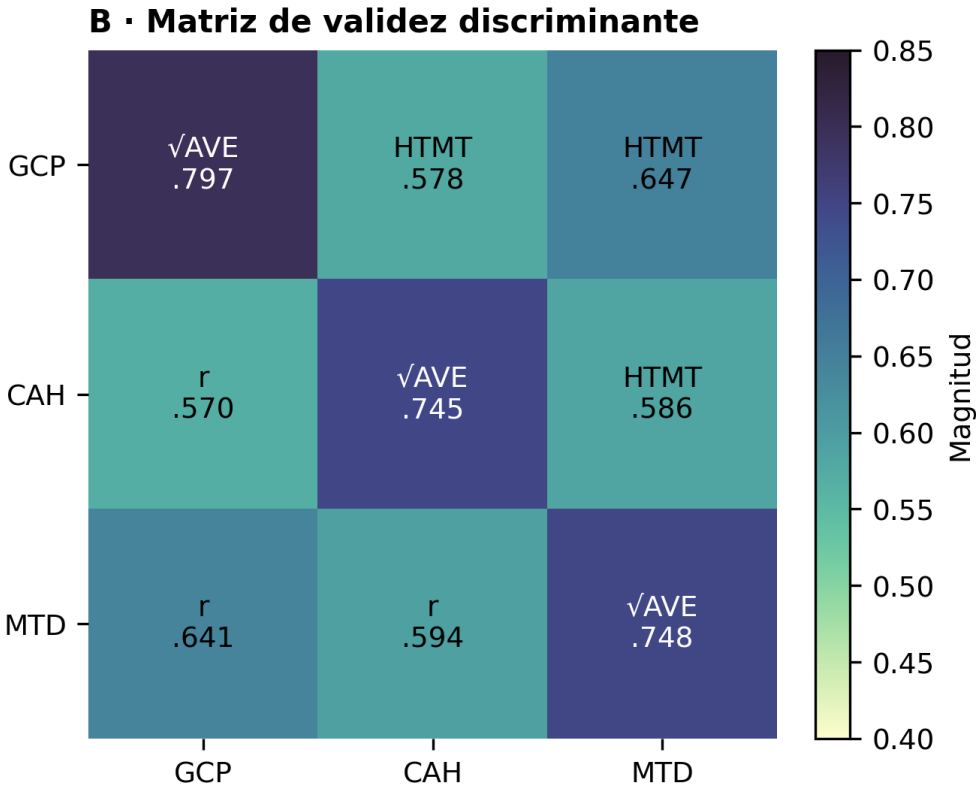
(ruta b). El efecto directo de la gestión sobre la calidad del aprendizaje se mantuvo positivo y significativo (c′), de modo que H1 resultó soportada (Tabla 3-6).

Figura 3-1

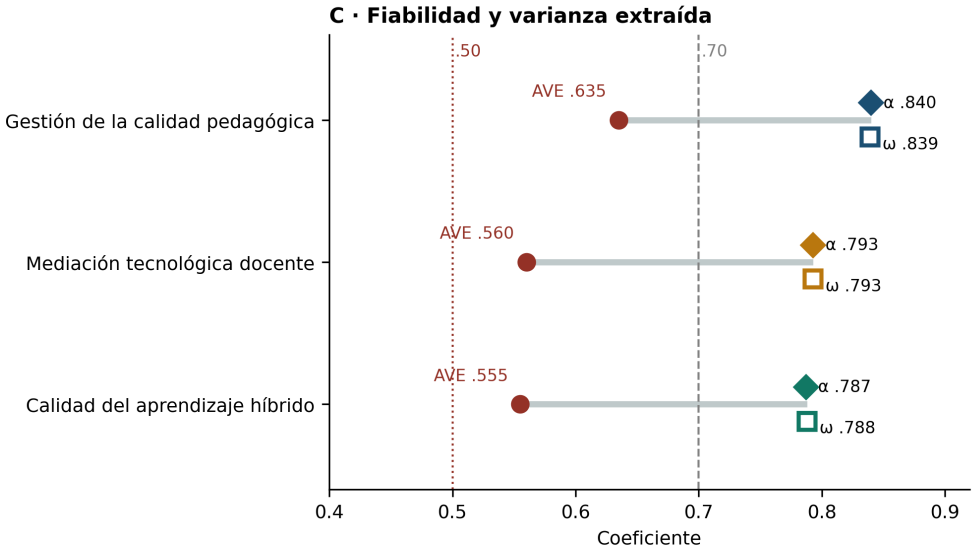
Modelo de medición, validez discriminante y ajuste de los modelos competidores.

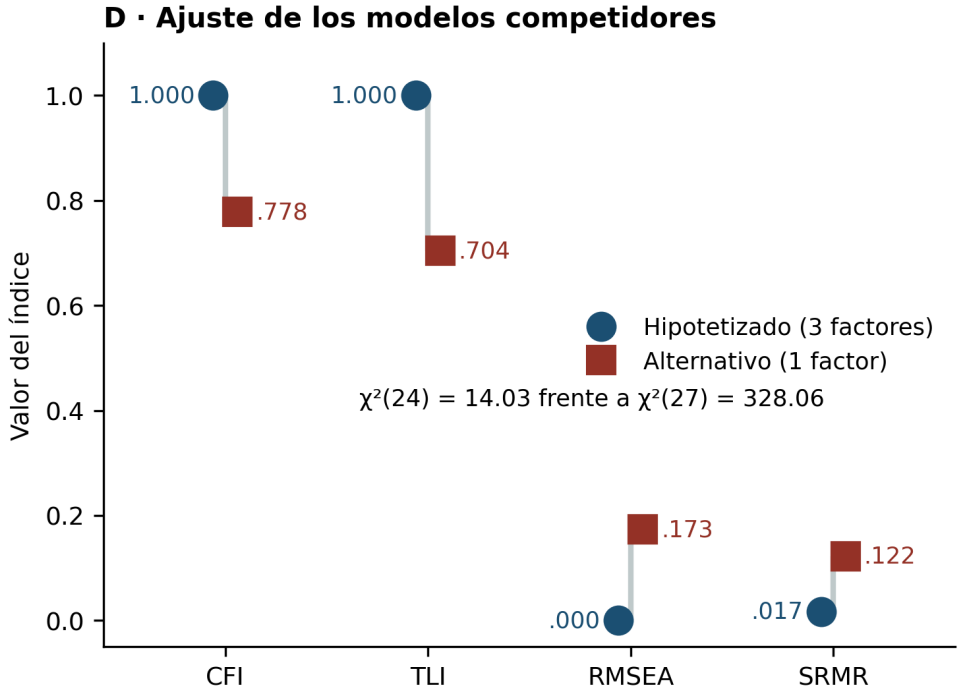


Nota. N = 375 docentes; análisis factorial confirmatorio por máxima verosimilitud. Panel A: cargas estandarizadas de los nueve ítems (rango .687–.831), con la referencia de .70 y la franja λ < .50 sombreada; códigos según la Tabla 3-1.



Nota. N = 375 docentes; análisis factorial confirmatorio por máxima verosimilitud. Panel B: matriz de validez discriminante con √AVE en la diagonal (GCP .797, CAH .745, MTD .748), correlaciones entre constructos en el triángulo inferior (r = .570, .641 y .594) y HTMT en el superior (.578, .647 y .586; criterio < .85).





Nota. Cálculo de potencia a priori y modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud (N = 375). Panel C: α de Cronbach, ω y AVE por constructo (GCP .840, .839 y .635; MTD .793, .793 y .560; CAH .787, .788 y .555), con las referencias .70 y .50. Panel D: CFI, TLI, RMSEA y SRMR del modelo hipotetizado de tres factores (1.000, 1.000, .000 y .017; $\chi^2(24) = 14.03$) frente al modelo alternativo de un factor (.778, .704, .173 y .122; $\chi^2(27) = 328.06$). GCP = gestión de la calidad pedagógica; MTD = mediación tecnológica docente; CAH = calidad del aprendizaje híbrido.

El intervalo bootstrap del efecto indirecto excluyó el cero, lo que respalda H2. Al ser también significativo el efecto directo, la mediación fue parcial: el efecto indirecto representó el 39.8 % del efecto total (.210 / .528) (Tabla 3-7).

Tabla 3-7

Descomposición de efectos de la mediación

Efecto		Coeficiente	IC 95 %		Significativo
			inferior	superior	
a	a (GCP → MTD)	.607	.500	.704	Sí
b	b (MTD → CAH)	.346	.191	.504	Sí
c′ directo	c′ directo (GCP → CAH)	.318	.167	.461	Sí
a × b	a × b indirecto	.210	.116	.322	Sí
c total	c total (c′ + a × b)	.528	.428	.620	Sí

Nota: Coeficientes del modelo estructural latente; intervalos al 95 % por bootstrap de 5000 réplicas (5000 convergentes); significativo si el intervalo excluye el cero.

3.4.4 Análisis multigrupo por zona

Las medias y varianzas de los tres constructos no difirieron significativamente entre zonas (p entre .103 y .848; Tabla 3-8).

Tabla 3-8

Comparación de los constructos entre zonas

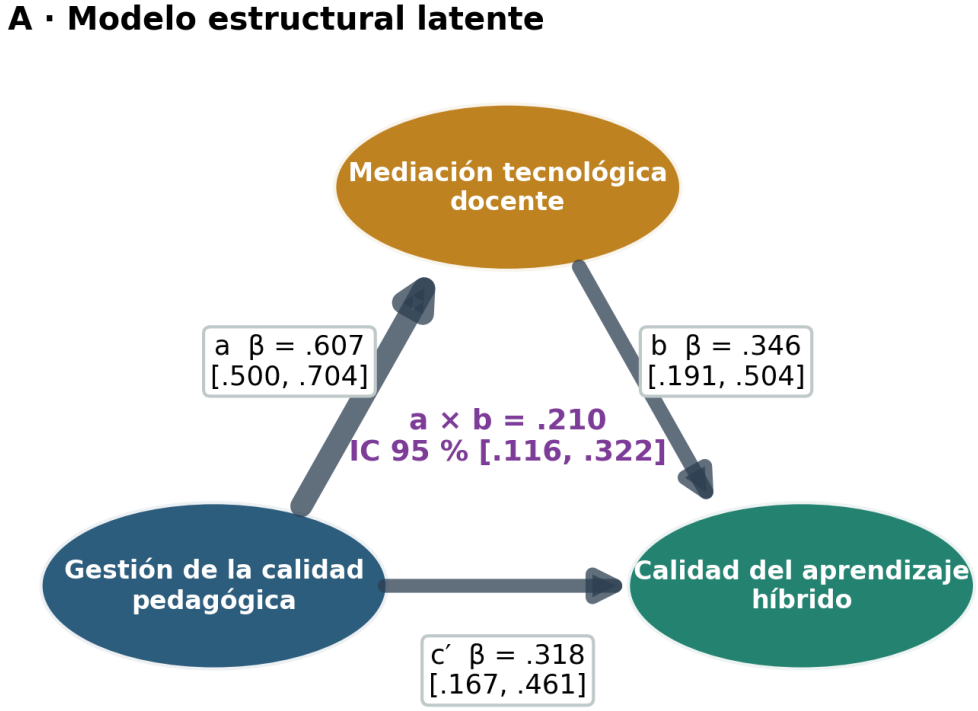
Constructo	p (igualdad de...	
	...medias, Welch)	...varianzas, Levene)
Gestión de la calidad pedagógica	.795	.666
Calidad del aprendizaje híbrido	.848	.244
Mediación tecnológica docente	.103	.248

Nota: Zona rural (n = 179) frente a zona urbana (n = 196). La invarianza de medición entre zonas se informa en la Tabla 3-5.

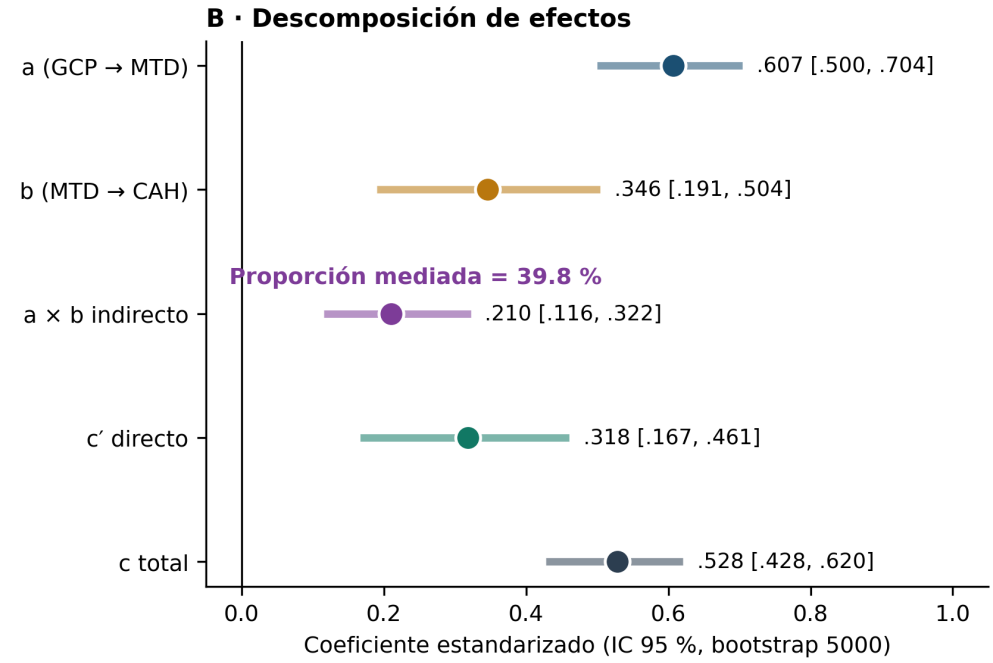
La pendiente de la gestión sobre la calidad del aprendizaje fue positiva y significativa en ambas zonas, y mayor en la urbana que en la rural; la diferencia fue significativa, por lo que H3 resultó soportada (Tabla 3-9).

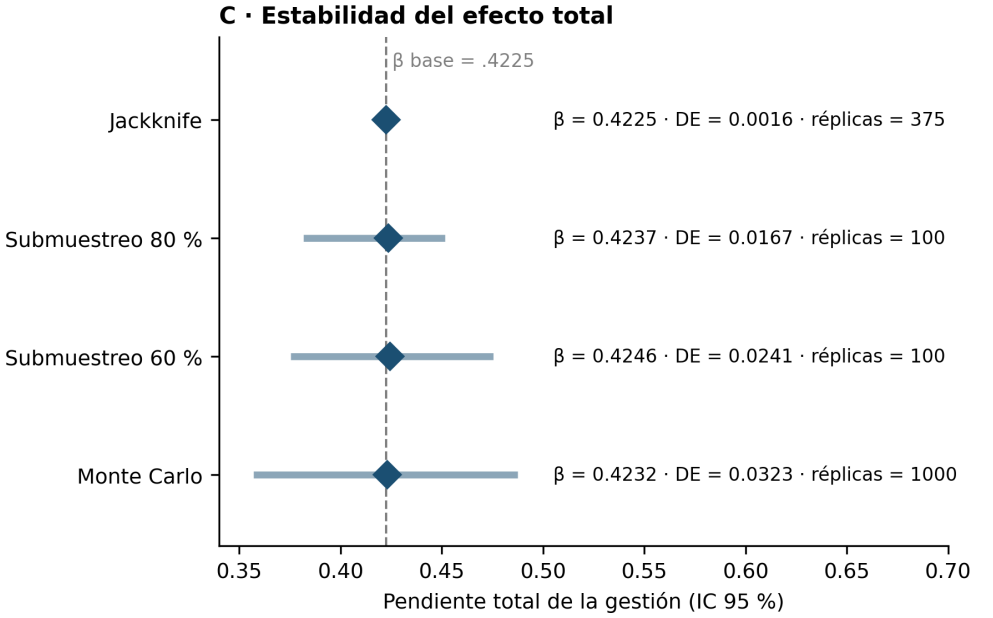
Figura 3-2

Mecanismo estructural, descomposición de efectos y robustez del efecto de la gestión

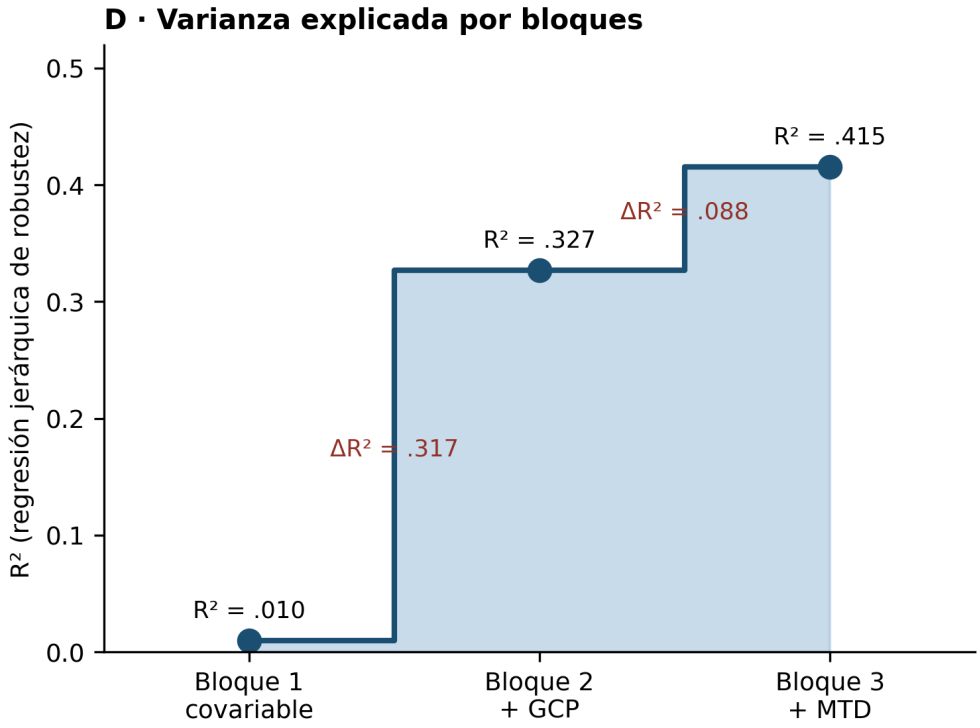


Nota. N = 375. Panel A: modelo estructural latente estimado por máxima verosimilitud, con el grosor de cada flecha proporcional a β y su IC 95 % bootstrap (5000 réplicas): a = .607 [.500, .704], b = .346 [.191, .504], c′ = .318 [.167, .461] y efecto indirecto a × b = .210 [.116, .322].





Nota. N = 375. Panel B: coeficientes e intervalos de las rutas a y b, del efecto indirecto, del efecto directo c′ y del efecto total c = .528 [.428, .620]; la proporción mediada es del 39.8 %. Panel C: pendiente total de la gestión sobre la calidad del aprendizaje estimada por mínimos cuadrados con las puntuaciones de los constructos (β base = .4225) en jackknife (375 réplicas), submuestreo al 80 % y al 60 % (100 réplicas cada uno) y Monte Carlo (1000 réplicas), con su IC 95 %; el 100 % de las réplicas conservó signo y significancia.



Nota. N = 375. Panel D: R^2 de la regresión jerárquica de robustez por bloques (covariable experiencia docente híbrida .010; + gestión .327; + mediación tecnológica .415), con ΔR^2 = .317 y .088. GCP = gestión de la calidad pedagógica; MTD = mediación tecnológica docente; CAH = calidad del aprendizaje híbrido.

Tabla 3-9

Análisis multigrupo del efecto de la gestión sobre la calidad del aprendizaje

Grupo	n	β	EE	p
Zona rural	179	.358	.044	< .001
Zona urbana	196	.482	.042	< .001
Diferencia Δβ (urbana – rural)	—	.123	—	.043

Nota: Pendiente de la regresión por mínimos cuadrados de la calidad del aprendizaje sobre la gestión, estimada con las puntuaciones de los constructos (efecto total, sin el mediador); por ello su magnitud se sitúa en torno al efecto total y no coincide con el efecto directo c′ de la Tabla 3-6. Δβ probado con el término grupo × gestión y errores robustos HC3.

3.4.5 Robustez

El efecto total de la gestión se mantuvo estable: el 100 % de las réplicas de cada prueba conservó el signo positivo y la significancia (Tabla 3-10).

Tabla 3-10

Estabilidad del efecto total de la gestión bajo perturbaciones de la muestra

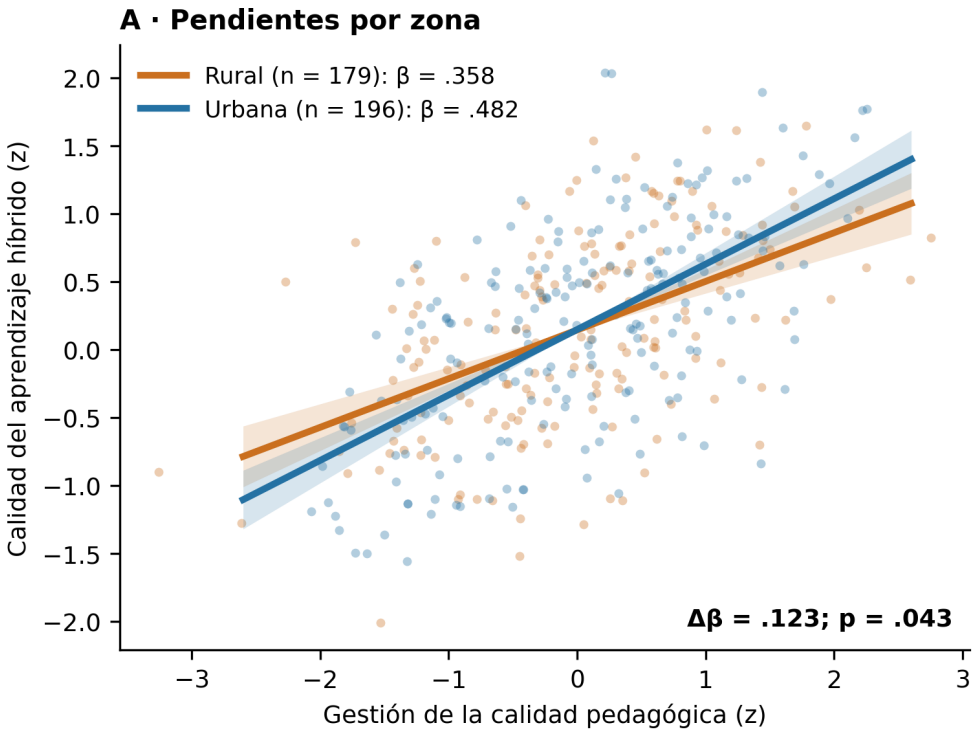
Prueba	Re-muestras	β medio	DE	iIC 95 %	Signo estable (%)	Significativo (%)
Jackknife (excluyendo un caso)	375	.4225	.0016	[.4191, .4259]	100 %	100 %
Submuestreo 80 %	100	.4237	.0167	[.3817, .4516]	100 %	100 %
Submuestreo 60 %	100	.4246	.0241	[.3755, .4755]	100 %	100 %
Monte Carlo (remuestreo con reemplazo)	1000	.4232	.0323	[.3571, .4875]	100 %	100 %

Nota: Pendiente por mínimos cuadrados de la calidad del aprendizaje sobre la gestión con las puntuaciones de los constructos (efecto total, β base = .4225), estimada en cada réplica; porcentajes de réplicas con el mismo signo y con p < .05.

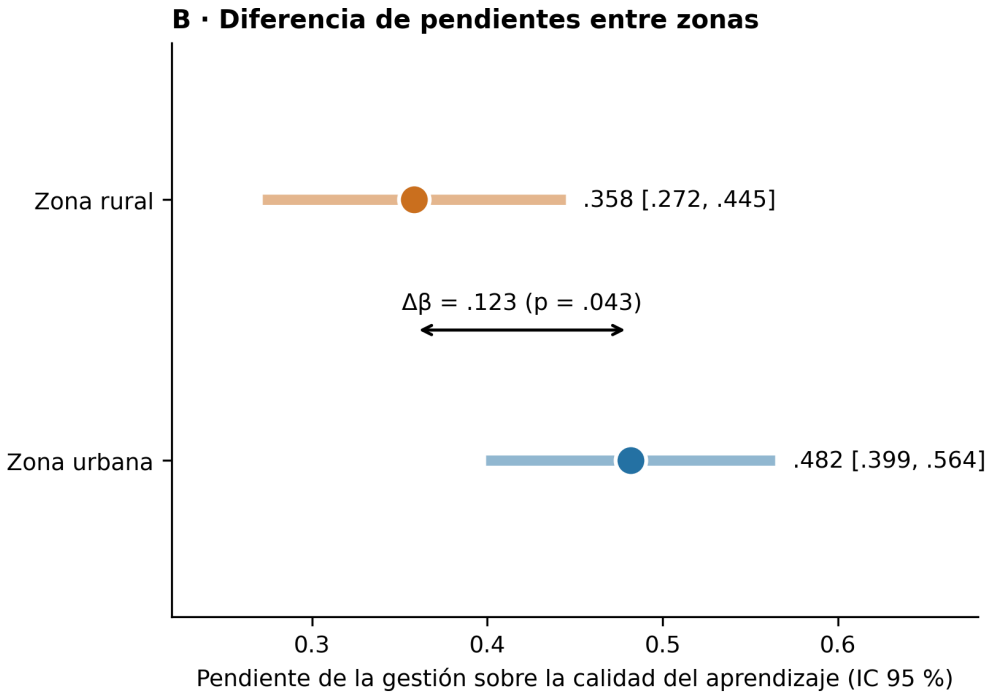
La regresión jerárquica replicó la dirección y la significancia de las rutas. La experiencia docente híbrida no fue significativa en ningún bloque (β = .015, p = .059 en el bloque de control; β = .007, p = .322 en el final).

Figura 3-3

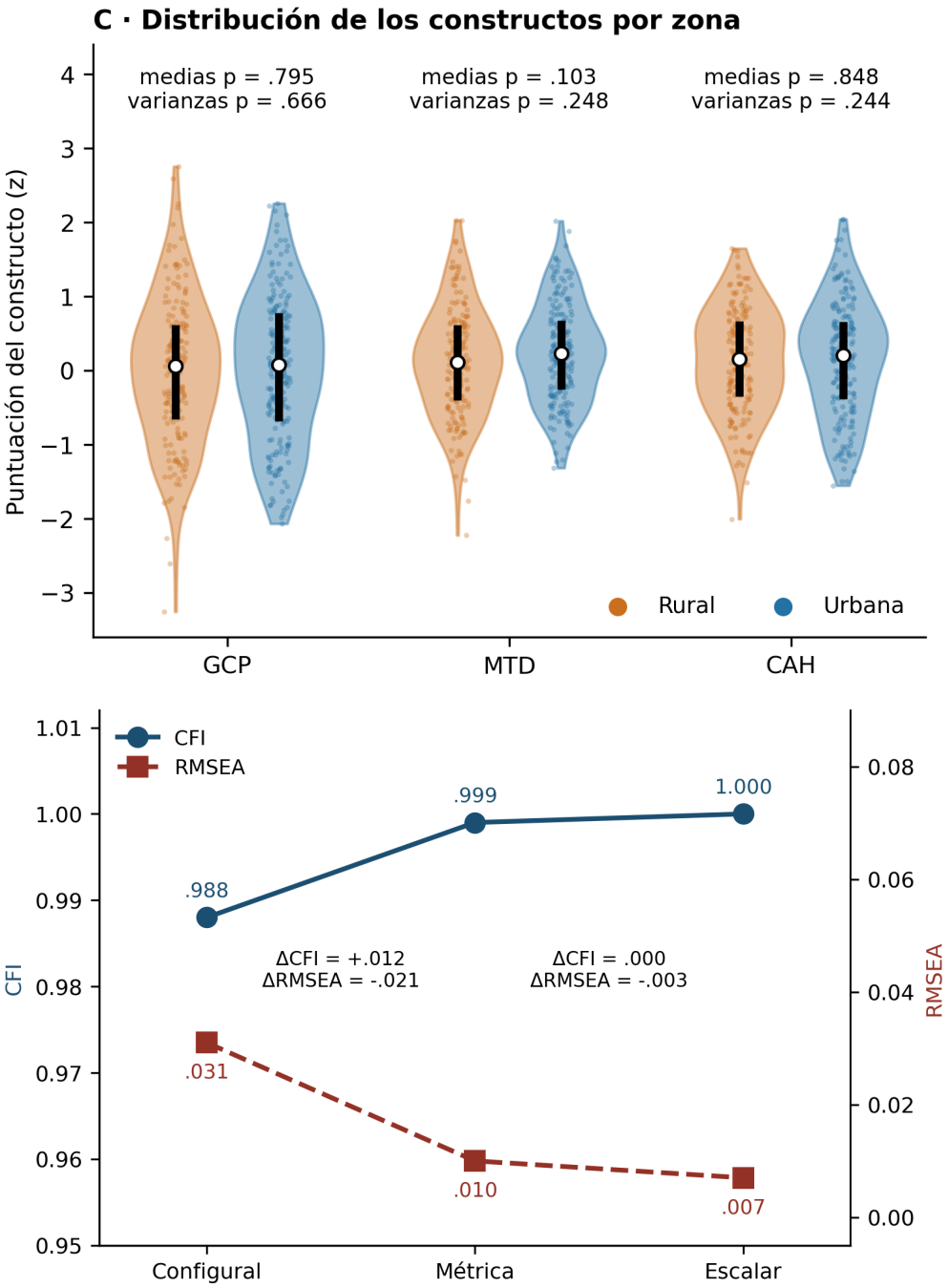
Contingencia territorial del efecto de la gestión de la calidad pedagógica



Nota. Zona rural n = 179; zona urbana n = 196. Panel A: dispersión de las puntuaciones estandarizadas de gestión y de calidad del aprendizaje por zona, con la pendiente de cada zona trazada por la media del grupo y su banda de incertidumbre (± 1.96 EE de la pendiente): rural β = .358, urbana β = .482; Δβ = .123, p = .043 (término grupo × gestión, errores HC3).



Nota. Zona rural n = 179; zona urbana n = 196. Panel B: pendientes por zona con IC 95 % calculado como β ± 1.96 EE (rural .358 [.272, .445]; urbana .482 [.399, .564]) y diferencia Δβ.



Nota. Zona rural n = 179; zona urbana n = 196. Panel C: distribución de los tres constructos por zona (violín, puntos individuales, mediana y rango intercuartílico) con las pruebas de igualdad de medias (Welch) y de varianzas (Levene): GCP p = .795 y .666; MTD p = .103 y .248; CAH p = .848 y .244. Panel D: CFI y RMSEA de los modelos de invarianza configural (.988 y .031), métrica (.999 y .010) y escalar (1.000 y .007), con ΔCFI = +.012 y .000 y ΔRMSEA = −.021 y −.003; el CFI no disminuye en ningún paso. GCP = gestión de la calidad pedagógica; MTD = mediación tecnológica docente; CAH = calidad del aprendizaje híbrido.

La gestión aportó un incremento de R^2 de .317 (β = .422, IC 95 % [.361, .484]) y la mediación tecnológica docente añadió .088 (β = .410, IC 95 % [.309, .511]), con la gestión aún significativa (β = .237, IC 95 % [.160, .314]), para un R^2 final de .415 y un VIF máximo de 1.01. Al no declararse moderadores, el bloque completo reproduce el último bloque y se informa por completitud.

3.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación positiva entre gestión y aprendizaje, que la literatura anticipaba, sino que los docentes urbanos y rurales no difieren en el nivel medio ni en la dispersión de ninguno de los tres constructos, interpretan los ítems de manera equivalente y, aun así, una misma intensidad de gestión parece traducirse en menos calidad del aprendizaje en las instituciones rurales. La diferencia no reside, por tanto, en cuánta gestión se declara ni en cómo se entiende, sino en su rendimiento. Conviene subrayar que la diferencia de pendientes es moderada, se sitúa cerca del umbral convencional de significancia y se estimó sobre el efecto total, sin el mediador; los datos sugieren una atenuación, no una ruptura, del mecanismo en el medio rural. Con esa cautela, el patrón respalda la idea de que la gestión no opera como una constante organizacional, sino como un parámetro condicionado por el entorno en que se ejerce la docencia.

Este resultado dialoga de forma matizada con la evidencia sobre liderazgo en contextos rurales. En escuelas rurales chinas, promover el aprendizaje docente ejerce una influencia dominante sobre la calidad de la instrucción, mientras que gestionar la instrucción tiene un efecto significativo pero débil (Lin y Hu, 2026). Si se acepta que el constructo de gestión aquí medido —planificación con estándares, monitoreo con retroalimentación y uso del monitoreo en ciclos de mejora— se aproxima más a la gestión de la instrucción que al desarrollo profesional, la atenuación rural observada es consistente con esa lectura; sin embargo, el diseño no permite distinguir dimensiones de la gestión y el contraste procede de un sistema educativo distinto, de modo que la analogía es sugerente y no concluyente. Desde una perspectiva crítica, tampoco cabe descartar que el menor rendimiento

rural obedezca a condiciones materiales no medidas, como la conectividad o el soporte técnico, pues el estudio no incorporó indicadores de acceso. La equivalencia de las medias de mediación tecnológica entre zonas indica que los docentes rurales declaran prácticas comparables, pero no que dispongan de las mismas condiciones para ejercerlas ni que obtengan de ellas el mismo resultado. En esta línea, la evidencia de que la escuela puede reducir o reproducir las desigualdades en habilidades digitales según sus condiciones de operación (Loh et al., 2025), y la identificación de la gestión educativa y la infraestructura tecnológica como factores institucionales de la competencia digital docente en América Latina (Ramírez et al., 2025), sugieren que la gestión y la infraestructura están entrelazadas. Desplazar el análisis de la brecha digital educativa desde el acceso hacia la eficacia de la gestión es, así, una hipótesis fértil y no una conclusión cerrada, que exigiría modelar conjuntamente ambos planos.

En cuanto al mecanismo, una parte sustantiva del efecto total transita por la mediación tecnológica docente y el efecto directo permanece. Este patrón se asemeja al de los modelos en los que el liderazgo escolar influye en las prácticas de enseñanza a través de la eficacia colectiva docente (Özdemir et al., 2025), de la autonomía y la innovatividad colectiva en serie (Demir y Kalman, 2025), de la autoeficacia y la cultura organizacional (Long y Xia, 2025) o de la autonomía y la innovación docentes en datos TALIS (Hsieh et al., 2025). La diferencia sustantiva radica en el mediador: aquí no se trata de una disposición psicosocial del profesorado, sino de un repertorio de prácticas propio de la docencia mediada por tecnología —seleccionar recursos según el propósito y la conectividad, andamiar en plataformas, adaptar el diseño instruccional—, de modo que el estudio extiende a la modalidad híbrida una lógica ya conocida en contextos presenciales. Que la porción restante del efecto no pase por este mediador sugiere que operan otros mecanismos no modelados. Por la naturaleza transversal de los datos, el efecto indirecto debe leerse como un patrón de asociación compatible con el mecanismo propuesto, no como una prueba causal.

La persistencia del efecto directo coincide con la evidencia de que el liderazgo curricular se asocia directa e indirectamente con la calidad de

la instrucción, y de que la agencia docente explica parte de esa asociación (Kuloğlu et al., 2026), así como con las prácticas directivas de liderazgo distribuido que promueven el desarrollo profesional docente en escuelas públicas chilenas (Galdames, 2023).

Por otra parte, la magnitud moderada de la ruta entre mediación tecnológica y aprendizaje invita a considerar los límites de la propia mediación. La competencia digital docente no es homogénea: la comparación entre docentes en formación y en servicio con el marco DigCompEdu documenta brechas que motivan propuestas para cerrarlas (Momdjian et al., 2024), y la mitad de los docentes en formación encuestados no se siente suficientemente preparada por su programa para fomentar la competencia digital (Dolezal et al., 2025). Si la preparación docente acota lo que la mediación puede transmitir, una gestión de calidad que no vaya acompañada de formación tendría un techo de rendimiento, y esa restricción podría ser más severa donde el apoyo es más escaso, como en las instituciones rurales; se trata de una conjetura que futuros estudios deberían contrastar con datos de formación y soporte. Un análisis de la gestión académica de la educación virtual universitaria en Perú, Colombia, México y España apunta en una dirección afín: el 51 % de los docentes está de acuerdo con la gestión y el 50 % acepta el uso adecuado de herramientas tecnológicas, pero solo el 20 % considera eficaces las clases virtuales (Salazar et al., 2023). La distancia entre la gestión percibida y la eficacia percibida es compatible con la idea de que la gestión rinde en la medida en que se convierte en práctica docente efectiva.

Las implicaciones prácticas siguen esa doble estructura. Puesto que parte del efecto de la gestión pasa por la mediación tecnológica, el aseguramiento de la calidad de la enseñanza híbrida debería incorporar indicadores sobre cómo el docente selecciona, adapta y acompaña con recursos digitales, y no solo sobre la disponibilidad de plataformas. Los equipos directivos podrían priorizar ciclos de monitoreo con retroalimentación centrados en la mediación, y la formación docente continua debería abordar de manera explícita el andamiaje y la adaptación a condiciones limitadas. Puesto que el rendimiento de la gestión es menor en zonas rurales, la

política educativa del Perú y de sistemas con desigualdades territoriales semejantes debería acompañar la inversión en conectividad con apoyo pedagógico diferenciado para las instituciones rurales, sin presumir que cerrar la brecha de acceso baste para cerrar la brecha de eficacia. Para la teoría de la gestión y el liderazgo pedagógico, el estudio sugiere incorporar la mediación tecnológica como eslabón de la cadena de conversión y el territorio como condición de contorno.

El estudio tiene limitaciones que acotan estas lecturas. El diseño transversal no establece precedencia temporal. Las medidas proceden del autorreporte docente; aunque el modelo de un factor descartó un factor común de método como explicación única de las covarianzas, convendría incorporar indicadores de aprendizaje de los estudiantes o de registros institucionales. Los índices de ajuste del modelo hipotetizado se sitúan en sus valores límite porque el χ^2 resultó inferior a los grados de libertad, por lo que conviene leerlos con prudencia y replicarlos en otras muestras. El muestreo por conveniencia impide generalizar a la población docente del sur peruano. El análisis no modeló el anidamiento de los docentes en instituciones, y la comparación entre zonas se basó en la diferencia de pendientes con errores robustos sobre constructos invariantes, no en un modelo latente multigrupo completo; además, esa pendiente corresponde al efecto total y no al efecto directo, y su diferencia es modesta. La experiencia docente híbrida no resultó significativa en la regresión jerárquica, de modo que no explica los patrones descritos, pero otras covariables institucionales quedaron sin controlar. Estudios longitudinales y multinivel, con medidas de conectividad y de formación docente, podrían contrastar si la brecha en el rendimiento de la gestión se reduce al cerrarse la brecha de acceso.

3.6 CONCLUSIONES

La gestión de la calidad pedagógica se asocia con la calidad del aprendizaje híbrido por una vía directa y por una vía indirecta a través de la mediación tecnológica docente, que transmite cerca del 40 % del efecto total. Las tres hipótesis recibieron respaldo empírico en una muestra de docentes de secundaria y universidad del sur peruano, con un modelo de medición

sólido, validez discriminante establecida e invarianza de medición entre zonas. Esa conversión no es territorialmente neutra: es más intensa en las instituciones urbanas que en las rurales, aun con constructos equivalentes y niveles medios comparables entre zonas. La brecha de mediación tecnológica se expresa así como una brecha en el rendimiento de la gestión pedagógica, y no únicamente como una diferencia en el acceso o en los niveles declarados de gestión.

Desde el punto de vista teórico, el estudio incorpora la mediación tecnológica docente como eslabón de la cadena que convierte la gestión institucional en aprendizaje híbrido y muestra que el territorio opera como condición de contorno de esa conversión. Desde el punto de vista práctico, exige que el aseguramiento de la calidad en la educación híbrida integre la mediación tecnológica del docente entre sus indicadores y que la política educativa combine la expansión de la conectividad con apoyo pedagógico diferenciado para los contextos rurales. Estas conclusiones deben leerse con la cautela que imponen el diseño transversal, el autorreporte y el muestreo por conveniencia. La agenda pendiente incluye diseños longitudinales y multinivel que incorporen medidas de conectividad, de formación docente y de aprendizaje de los estudiantes, para determinar si la brecha de eficacia de la gestión persiste cuando se equipara el acceso.

3.7 REFERENCIAS BIBLIOGRÁFICAS

Adigun, O., Mpofu, N., & Maphalala, M. (2025). Fostering self-directed learning in blended learning environments: A constructivist perspective in Higher Education. *Higher Education Quarterly*, *79*(1). https://doi.org/10.1111/hequ.12572

Agasisti, T., Antequera, G., & Delprato, M. (2023). Technological resources, ICT use and schools efficiency in Latin America – Insights from OECD PISA 2018. *International Journal of Educational Development*, *99*, 102757. https://doi.org/10.1016/j.ijedudev.2023.102757

Ancheta, A., Pulido, C., & Carvajal, V. (2021). Gender Digital Divide and Education in Latin America: A Literature Review. *Education Sciences*, *11*(12), 804. https://doi.org/10.3390/educsci11120804

Batista, S., & Gavilan, D. (2023). Student Experience, Satisfaction and Commitment in Blended Learning: A Structural Equation Modelling Approach. *Mathematics*, *11*(3), 749. https://doi.org/10.3390/math11030749

Bertoni, E., Elacqua, G., Hincapié, D., Méndez, C., & Paredes, D. (2023). Teachers' Preferences for Proximity and the Implications for Staffing Schools: Evidence from Peru. *Education Finance and Policy*, *18*(2), 181-212. https://doi.org/10.1162/edfp_a_00347

Bonifaz, J., & Aguirre, J. (2026). Impact assessment of broadband transport and internet access projects in Peru. *Telecommunications Policy*, *50*(3), 103149. https://doi.org/10.1016/j.telpol.2025.103149

Borchers, J., & Cunha, M. (2025). School performance and inequality of opportunities in Latin America. *Studies in Educational Evaluation*, *86*, 101497. https://doi.org/10.1016/j.stueduc.2025.101497

Brenya, B. (2024). Higher education in emergency situation: blended learning prospects and challenges for educators in the developing countries. *Journal of Applied Research in Higher Education*, *16*(4), 1015-1028. https://doi.org/10.1108/jarhe-01-2023-0044

Califano, B., & Becerra, M. (2021). Public policies of connectivity in Latin America in the context of COVID-19. *Journal of Digital Media & Policy*, *12*(1), 117-136. https://doi.org/10.1386/jdmp_00051_1

Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Clifft, S., & Assiouras, I. (2023). The transformation of post pandemic hybrid teaching and learning through experiences of remote digital learning in French business schools. *Innovations in Education and Teaching International*, *60*(6), 810-823. https://doi.org/10.1080/14703297.2022.2132982

Concha, M., Serpa, A., Vivas, L., & Saintila, J. (2025). Impact of Online Learning Strategies on Academic Engagement in Peruvian University Students: Moderating Role of Emotional Exhaustion Levels. *Electronic Journal of Research in Education Psychology*, *23*(67), 505-526. https://doi.org/10.25115/ejrep.v23i67.10104

De la Portilla, J. (2023). La estrategia educativa 'Aprendo en casa' y la implementación del enfoque intercultural. *Investigaciones Sociales*(48), 173-190. https://doi.org/10.15381/is.n48.24135

Demir, K., & Kalman, M. (2025). Instructional Leadership and Teachers' Instructional Practices: A Serial Mediation by Teacher Autonomy and Collective Teacher Innovativeness. *European Journal of Education*, *60*(3). https://doi.org/10.1111/ejed.70155

Dolezal, D., Motschnig, R., & Ambros, R. (2025). Pre-Service Teachers' Digital Competence: A Call for Action. *Education Sciences*, *15*(2), 160. https://doi.org/10.3390/educsci15020160

Galdames, M. (2023). Distributed Leadership: School Principals' Practices to Promote Teachers' Professional Development for School Improvement. *Education Sciences*, *13*(7), 715. https://doi.org/10.3390/educsci13070715

Greere, A., & Crozier, F. (2024). Quality assurance expectations for online higher education: stepping stones to support post-pandemic decisions in Georgia. *Quality in Higher Education*, *30*(1), 29-54. https://doi.org/10.1080/13538322.2022.2123266

Hafeez, M., Naureen, S., & Sultan, S. (2022). Quality Indicators and Models for Online Learning Quality Assurance in Higher Education. *Electronic Journal of e-Learning*, *20*(4), 374-385. https://doi.org/10.34190/ejel.20.4.2553

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hsieh, C., Song, Y., & Li, H. (2025). Analyzing the relationship between distributed leadership and instructional quality in Taiwan: The mediating roles of teacher autonomy and teacher innovation. *Educational Management Administration & Leadership*, *53*(5), 1153-1171. https://doi.org/10.1177/17411432241231421

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Justo, J., Chambi, E., Vilca, J., Diaz, Y., & Atencio, N. (2025). Quality and satisfaction of virtual education in students of a public university in Peru. *International Journal of Innovative Research and Scientific Studies*, *8*(1), 2001-2008. https://doi.org/10.53894/ijirss.v8i1.4872

Kuloğlu, A., Alanoglu, M., & Karabatak, S. (2026). Curriculum leadership and instructional quality: A moderated mediation model of teacher agency and leader–member exchange. *Educational Management Administration & Leadership*. https://doi.org/10.1177/17411432261490287

Lin, Y., & Hu, D. (2026). Instructional leadership malfunction in rural areas: Examining the influence of principal instructional leadership on instructional quality in Chinese rural schools. *Educational Management Administration & Leadership*, *54*(5), 1572-1589. https://doi.org/10.1177/17411432251346085

Loh, R., Kraaykamp, G., & Van Hek, M. (2025). Plugging in at school: Do schools nurture digital skills and narrow digital skills inequality?. *Computers & Education*, *226*, 105195. https://doi.org/10.1016/j.compedu.2024.105195

Long, Y., & Xia, Y. (2025). Leveraging transformational and instructional leadership for teacher professional development: A dual-mediation model of teacher self-efficacy and organisational culture. *Teaching and Teacher Education*, *164*, 105087. https://doi.org/10.1016/j.tate.2025.105087

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling.

Psychological Methods, 1(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Momdjian, L., Manegre, M., & Gutiérrez, M. (2024). Assessing and bridging the digital competence gap: a comparative study of lebanese student teachers and in-service teachers using the DigCompEdu framework. *Discover Education, 3*(1). https://doi.org/10.1007/s44217-024-00308-2

Özdemir, S., Sezgin, F., Kilinç, A., & Polatcan, M. (2025). A cultural lens to school leadership effects on teacher instructional practices: The mediation of teacher collective efficacy and the moderation of uncertainty avoidance. *Educational Management Administration & Leadership, 53*(3), 515-535. https://doi.org/10.1177/17411432231177536

Panigrahi, R., Nihar, K., & Singh, N. (2024). Quality Measurement of Blended Learning Model in Higher Education: Scale Development and Validation. *Higher Learning Research Communications, 14*(1). https://doi.org/10.18870/hlrc.v14i1.1467

Pérez, G., Amaya, R., Arhuis, W., Tamayo, C., & Valverde, B. (2026). Digital competence and learning strategies in preschool teachers in Peru. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1686920

Quast, J., Rubach, C., & Porsch, R. (2025). Professional digital competence beliefs of student teachers, pre-service teachers and teachers: Validating an instrument based on the DigCompEdu framework. *European Journal of Teacher Education, 48*(4), 698-721. https://doi.org/10.1080/02619768.2023.2251663

Ramírez, R., Franco, E., & Carcausto, W. (2025). Factors Influencing the Development of Digital Competencies in Higher Education Teachers in Latin America: A Systematic Review. *Journal of Educational and Social Research*, *15*(1), 104. https://doi.org/10.36941/jesr-2025-0009

Remesal, A., & Villarroel, V. (2023). Challenges for Post-Pandemic Virtual Education in Latin America: A Comparative Analysis of the Emergency Remote Higher Education Process in Chile, Mexico, and Ecuador. *Sustainability*, *15*(19), 14199. https://doi.org/10.3390/su151914199

Rubio, M., Cabero, J., & Palacios, A. (2023). Digital Innovation in Language Teaching—Analysis of the Digital Competence of Teachers according to the DigCompEdu Framework. *Education Sciences*, *13*(4), 336. https://doi.org/10.3390/educsci13040336

Salazar, C., Zegarra, M., & Cordova, F. (2023). Academic management of university virtual education: An analysis from the perception of students, teachers, and managers. *Problems and Perspectives in Management, 21*(4), 532-544. https://doi.org/10.21511/ppm.21(4).2023.40

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13-35. https://doi.org/10.1016/j.compedu.2018.09.009

CAPÍTULO IV

GOBERNANZA INSTITUCIONAL MEDIANTE COMPROMISOS DE GESTIÓN ESCOLAR Y RENDIMIENTO ACADÉMICO: EL PAPEL MEDIADOR DE LA PERMANENCIA ESCOLAR EN LA SECUNDARIA PÚBLICA DEL PERÚ, 2025

Luis Enrique Rivas Pacherrez

Universidad Nacional de Piura

Ana Consuelo Yovera Espinoza

Universidad Nacional de Piura

Yonny Marvin Navarro Yanayaco

Universidad Nacional de Piura

Noelia Elizabeth Palacios Lopez

Universidad César Vallejo

Resumen

La gestión escolar por compromisos supone que las condiciones institucionales bien gobernadas sostienen la permanencia del estudiante y, a través de ella, su logro académico. Este estudio examinó si la gobernanza institucional, organizada en torno a los Compromisos de Gestión Escolar, se asocia con el rendimiento académico en la secundaria pública del Perú y si la permanencia escolar transmite parte de esa asociación. Se realizó un estudio cuantitativo, transversal y explicativo con 500 estudiantes de tercero a quinto grado, evaluados en 2025 con tres escalas reflectivas de tres ítems. Se estimó un modelo de ecuaciones estructurales basado en covarianzas por máxima verosimilitud, con intervalos de confianza bootstrap de 5000 réplicas para los efectos indirectos y un análisis de sensibilidad del estimador sobre puntuaciones compuestas. El modelo de medida mostró validez convergente y validez discriminante según la razón heterorrasgo-monorrasgo, y el modelo estructural ajustó a los datos. La gobernanza se asoció con la permanencia ($\beta = .824$) y con el rendimiento (efecto directo $\beta = .605$; efecto total $\beta = .771$). El efecto indirecto fue positivo ($\beta = .166$), pero su intervalo incluyó el cero, IC95% [−.032, .336]; con puntuaciones compuestas resultó distinto de cero ($\beta = .246$). La gobernanza escolar es un correlato robusto del rendimiento académico; la permanencia es un mecanismo plausible pero no concluyente cuando se corrige el error de medida, lo que orienta la evaluación de estas políticas hacia diseños longitudinales.

Palabras clave: gobernanza escolar; compromisos de gestión escolar; permanencia escolar; rendimiento académico; ecuaciones estructurales; educación secundaria.

Abstract

School management by commitments assumes that well-governed institutional conditions sustain student persistence and, through it, academic achievement. This study examined whether institutional governance organized around the

School Management Commitments is associated with academic achievement in Peruvian public secondary schools, and whether school persistence transmits part of that association. A quantitative, cross-sectional, explanatory study was conducted with 500 students in grades nine to eleven, assessed in 2025 with three reflective three-item scales. A covariance-based structural equation model was estimated by maximum likelihood, with 5000-replicate bootstrap confidence intervals for indirect effects and an estimator sensitivity analysis on composite scores. The measurement model showed convergent validity and discriminant validity according to the heterotrait-monotrait ratio, and the structural model fitted the data. Governance was associated with persistence (β = .824) and with achievement (direct effect β = .605; total effect β = .771). The indirect effect was positive (β = .166), but its interval included zero, 95% CI [−.032, .336]; with composite scores it differed from zero (β = .246). School governance is a robust correlate of academic achievement; persistence is a plausible but inconclusive mechanism once measurement error is corrected, which points the evaluation of these policies toward longitudinal designs.

Keywords: *school governance; school management commitments; school persistence; academic achievement; structural equation modeling; secondary education.*

4.1 INTRODUCCIÓN

La interrupción de las trayectorias escolares en la educación secundaria sigue siendo uno de los problemas más persistentes de los sistemas educativos latinoamericanos. Los análisis de cohortes muestran que una proporción considerable de jóvenes queda fuera de la escuela y del empleo en la transición a la adultez, con efectos acumulativos sobre la desigualdad (Székely & Karver, 2021), y las brechas de desempeño entre escuelas de la región se asocian con desigualdades de oportunidades que el propio sistema reproduce (Borchers & Cunha, 2025). El riesgo de abandonar la secundaria responde a factores comunes y a otros diferenciados por sexo (Adrogué & Orlicki, 2025), y las comunidades educativas atribuyen a la finalización de la secundaria un valor que trasciende la credencial:

desarrolla habilidades, amplía las oportunidades laborales y el acceso a la educación superior, y sustenta un proyecto de vida (Mendoza & Ballesta, 2023). En el Perú, los estudios sobre deserción identifican determinantes socioeconómicos, personales e institucionales (Calero et al., 2023), y los modelos predictivos a gran escala señalan la conducta y el rendimiento académico entre los factores que más contribuyen al riesgo de abandono (Barboza et al., 2026). Permanencia y logro aparecen, así, estrechamente entrelazados.

En ese marco, la política educativa peruana ha organizado la gestión de las instituciones públicas mediante los Compromisos de Gestión Escolar, que distinguen compromisos de resultados —el progreso de los aprendizajes y la permanencia de los estudiantes— de compromisos de condiciones de funcionamiento: la gestión del calendario y del tiempo lectivo, el acompañamiento y monitoreo de la práctica pedagógica y la gestión de la convivencia. Esta arquitectura encierra una hipótesis causal implícita: las condiciones institucionales bien gobernadas sostienen la permanencia del estudiante y, a través de ella, el logro académico. Los estudios nacionales disponibles han documentado el cumplimiento de estos compromisos desde la perspectiva de los directivos, con una correlación elevada respecto de la administración de recursos (Gómez & Coronel, 2023); han descrito la traducción de los lineamientos de política en prácticas pedagógicas y motivación estudiantil en escuelas públicas de Lima (López, 2026) y la relevancia de una gestión participativa y contextualizada en el sur del país (Solís et al., 2024); han revisado la gestión de la convivencia en las instituciones educativas peruanas (Magallanes et al., 2024) o han vinculado el desempeño docente con los logros de aprendizaje en la secundaria rural (Chocce, 2026). Ninguno, sin embargo, ha contrastado el mecanismo completo desde la experiencia de los estudiantes.

La evidencia internacional sustenta cada eslabón por separado. Un marco de liderazgo centrado en el aprendizaje predice el logro estudiantil mediante efectos directos e indirectos de magnitud comparable (Wu & Shen, 2025), y la relación entre liderazgo directivo y logro difiere entre sistemas educativos (Chen & Bos, 2023). La pertenencia escolar, núcleo psicológico

de la permanencia, media efectos longitudinales sobre el rendimiento (Song et al., 2024); el apoyo docente se relaciona longitudinalmente con el logro, y se distingue entre las asociaciones entre estudiantes y las que ocurren dentro de cada estudiante (Zhou et al., 2024); y ese apoyo opera a través de la autoeficacia y el compromiso (Huang & Wang, 2023). En Colombia, asistencia escolar y rendimiento se relacionan con las mismas condiciones materiales del hogar, lo que invita a tratarlos como resultados vinculados (Cabra, 2023).

Persiste, no obstante, un vacío en la articulación de estos hallazgos. Los modelos que vinculan liderazgo y logro rara vez incorporan la permanencia como mediador, pese a que la retención es, en la propia política peruana, un resultado intermedio con meta propia, y la mayoría de los estudios nacionales recurre a informantes directivos o a descripciones de estrategias, no a modelos que corrijan el error de medida. Este estudio aborda ese vacío con una contribución que trasciende el caso nacional: contrasta, con ecuaciones estructurales que corrigen el error de medida, la cadena gobernanza–permanencia–rendimiento implícita en los sistemas de gestión por compromisos y examina la sensibilidad de ese mecanismo al estimador. El Perú opera como escenario de prueba de un modelo de gestión escolar por resultados presente en numerosos sistemas educativos.

La pregunta de investigación es: ¿en qué medida la gobernanza institucional basada en los Compromisos de Gestión Escolar se asocia con el rendimiento académico de los estudiantes de secundaria pública del Perú, y en qué medida la permanencia escolar transmite esa asociación? A partir de ella se contrastan tres hipótesis, formuladas en la sección siguiente: un efecto directo de la gobernanza sobre el rendimiento (H1) y un efecto indirecto a través de la permanencia, contrastado por la significación conjunta de sus dos rutas (H2) y por el intervalo bootstrap del efecto indirecto (H3).

4.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

El modelo se apoya en la teoría de la eficacia escolar, que concibe la calidad de la gestión institucional como un conjunto de condiciones organizativas

que influyen en el aprendizaje a través de procesos intermedios, tal como lo formaliza el marco de liderazgo centrado en el aprendizaje (Wu & Shen, 2025). La gobernanza institucional se entiende aquí como la capacidad de la escuela para ordenar el tiempo lectivo, hacer seguimiento al aprendizaje y aplicar normas de convivencia de forma justa y oportuna: las tres condiciones de funcionamiento que los Compromisos de Gestión Escolar vuelven exigibles. Desde la perspectiva del estudiante, estas condiciones se perciben como previsibilidad, acompañamiento y equidad en el trato institucional.

La segunda base teórica procede de los modelos de compromiso y pertenencia escolar, según los cuales la permanencia no es solo un hecho administrativo sino una disposición sostenida del estudiante a continuar y a participar regularmente en la escuela. La pertenencia escolar media efectos sobre el rendimiento a lo largo del tiempo (Song et al., 2024), y la asistencia y el rendimiento se asocian con las mismas condiciones materiales del hogar (Cabra, 2023), lo que justifica tratarlos como resultados vinculados. El apoyo percibido de la institución y de los docentes fortalece esa disposición y, por esa vía, el desempeño: el apoyo docente actúa sobre el rendimiento mediante la autoeficacia y el compromiso (Huang & Wang, 2023) y mantiene relaciones longitudinales con el logro (Zhou et al., 2024).

De la integración de ambos marcos se derivan tres hipótesis. Si la gobernanza ordena las condiciones de enseñanza, debe asociarse directamente con el rendimiento académico (H1). Si, además, esas condiciones sostienen la disposición del estudiante a permanecer, la gobernanza debe vincularse con la permanencia y esta, a su vez, con el rendimiento una vez considerada la gobernanza (H2). Por último, si la permanencia es un mecanismo y no solo un correlato, el efecto indirecto de la gobernanza sobre el rendimiento a través de ella debe ser distinto de cero (H3).

- **H1.** La gobernanza escolar se asocia positivamente con el rendimiento académico (efecto directo).

- **H2.** Las dos rutas del mecanismo son positivas y distintas de cero: la gobernanza escolar se asocia con la permanencia escolar (ruta a) y la permanencia escolar se asocia con el rendimiento académico una vez considerada la gobernanza (ruta b) (criterio: significación conjunta de a y b, $p < .05$ en ambas).
- **H3.** La permanencia escolar media la asociación entre la gobernanza escolar y el rendimiento académico (criterio: el intervalo bootstrap percentil del efecto indirecto a × b excluye el cero).

4.3. MATERIALES Y MÉTODOS

4.3.1 Diseño

Se empleó un diseño cuantitativo, no experimental, transversal y explicativo, con un modelo de ecuaciones estructurales basado en covarianzas que especifica la gobernanza escolar como variable exógena, la permanencia escolar como mediadora y el rendimiento académico como variable de resultado. El nivel de análisis es el estudiante; el modelo es de un solo nivel.

4.3.2 Instrumentos

Los tres constructos se midieron con escalas reflectivas de tres ítems (nueve en total), de elaboración propia de los autores, con respuesta Likert de cinco puntos (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo). La gobernanza escolar se evaluó como percepción de las tres condiciones de funcionamiento de los Compromisos de Gestión Escolar: cumplimiento del calendario y del tiempo lectivo, seguimiento del aprendizaje y aplicación justa de las normas de convivencia. La permanencia escolar se evaluó como intención de permanencia, y el rendimiento académico, como rendimiento autopercibido en Comunicación, en Matemática y en comparación con los compañeros. La Tabla 4-1 presenta los ítems y las propiedades psicométricas del análisis factorial confirmatorio.

Tabla 4-1

Ítems del instrumento y propiedades del modelo de medida

Constructo	Ítem	λ	R^2	α	ω	AVE
Gobernanza escolar	En mi escuela se cumplen el calendario escolar y las horas de clase previstas	.755	.570	.808	.808	.584
	En mi escuela se hace seguimiento a mi aprendizaje y se me informa de mi progreso	.723	.523	.808	.808	.584
	En mi escuela las normas de convivencia se aplican de forma justa y oportuna	.812	.659	.808	.808	.584
Permanencia escolar	Tengo la intención de terminar la secundaria en esta escuela	.692	.479	.765	.765	.520
	Asisto a clases de forma regular durante todo el año escolar	.724	.524	.765	.765	.520
	Este año no he pensado en dejar los estudios	.747	.558	.765	.765	.520
Rendimiento académico	Mis calificaciones en Comunicación son buenas	.711	.506	.810	.809	.587
	Mis calificaciones en Matemática son buenas	.801	.642	.810	.809	.587
	En general, mi rendimiento académico es alto en comparación con el de mis compañeros	.783	.613	.810	.809	.587

Nota: λ = carga factorial estandarizada; R^2 = correlación múltiple al cuadrado del ítem; α = alfa de Cronbach; ω = omega de McDonald; AVE = varianza media extraída. Escala de respuesta Likert de 1 a 5. Criterios de referencia: λ ≥ .50, α y ω ≥ .70, AVE ≥ .50.

4.3.3 Participantes y tamaño de la muestra

Participaron estudiantes de tercero a quinto grado de secundaria de instituciones educativas públicas del Perú, evaluados entre abril y agosto de 2025. Se incluyó a quienes estaban matriculados en el año lectivo y completaron los nueve ítems; se excluyó a quienes dejaron respuestas incompletas. El muestreo fue no probabilístico por conveniencia, determinado por la accesibilidad de las instituciones que autorizaron la aplicación, por lo que los resultados describen a los estudiantes analizados y no se generalizan a la población estudiantil del país.

El tamaño muestral, 500 estudiantes, se fijó con un cálculo a priori que cubre el criterio más exigente (Tabla 2). Para el ajuste global se aplicó la potencia del ajuste cercano de MacCallum et al. (1996), que contrasta $\varepsilon 0 = .05$ frente a $\varepsilon 1 = .08$ con gl = 24 (9 indicadores y 21 parámetros libres). Para los coeficientes estructurales se empleó la prueba F de la regresión múltiple con un efecto mínimo de interés $f^2 = .02$ ($r \approx .14$), en el límite inferior del rango de las síntesis sobre liderazgo y clima escolar: efectos de 0.08 a 0.16 desviaciones estándar (Liebowitz & Porter, 2019), r = .178 (Demirtas et al., 2021) y la magnitud pequeña a media del apoyo docente percibido sobre el logro (Tao et al., 2022).

Tabla 4-2

Cálculo a priori del tamaño de la muestra

Criterio	Parámetros	N mínimo	Potencia con N = 500
Potencia del ajuste cercano (RMSEA)	gl = 24; $\varepsilon 0 = .05$; $\varepsilon 1 = .08$; $\alpha = .05$; $1 - \beta = .90$	492	.905
Potencia del ajuste cercano (RMSEA)	gl = 24; $\varepsilon 0 = .05$; $\varepsilon 1 = .08$; $\alpha = .05$; $1 - \beta = .80$	375	.905
Efecto parcial en la ecuación del rendimiento (2 predictores)	$f^2 = .02$; u = 2; $\alpha = .05$; $1 - \beta = .80$	485	.813
Efecto en la ecuación de la permanencia (1 predictor)	$f^2 = .02$; u = 1; $\alpha = .05$; $1 - \beta = .80$	395	.884

Nota: Potencia del ajuste cercano según MacCallum et al. (1996): $\lambda 0 = (N - 1) \cdot gl \cdot \varepsilon 0^2$, $\lambda 1 = (N - 1) \cdot gl \cdot \varepsilon 1^2$, potencia = $1 - F_\chi^2(c; gl, \lambda 1)$, con c = cuantil $1 - \alpha$ de la χ^2 no central de parámetro $\lambda 0$. Potencia de la prueba F: $\lambda = f^2 \cdot N$, potencia = $1 - F_F(c; u, N - u - 1, \lambda)$, con c = cuantil $1 - \alpha$ de la F central. El mínimo que satisface simultáneamente los cuatro criterios es 492; el N del estudio (500) lo supera en 8 casos y alcanza en ese criterio una potencia de .905.

4.3.4 Procedimiento

Los cuestionarios se aplicaron en las aulas, en horario escolar y en formato de lápiz y papel, durante unos quince minutos. Los estudiantes respondieron de forma anónima, con instrucciones que subrayaban la ausencia de respuestas correctas o incorrectas.

4.3.4 Análisis de datos

El análisis tuvo cuatro etapas. Primero, se estimaron descriptivos, asimetría, curtosis y correlaciones entre puntuaciones compuestas, con intervalos de confianza al 95 % por la transformación Z de Fisher. Segundo, se evaluó el modelo de medida mediante análisis factorial confirmatorio (cargas, α, ω y AVE) y la validez discriminante con el criterio de Fornell y Larcker (1981) y con la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015), con umbrales de .85 (estricto) y .90 (liberal). Tercero, se estimó el modelo estructural por máxima verosimilitud con la librería semopy de Python, y su ajuste se evaluó con χ^2, CFI, TLI, RMSEA y SRMR según los puntos de corte de Hu y Bentler (1999): CFI y TLI $\geq .95$, RMSEA $\leq .06$ y SRMR $\leq .08$. El modelo hipotetizado se comparó con uno alternativo de un solo factor mediante la diferencia de χ^2. Los coeficientes se reportan estandarizados, con errores estándar y valores p de la estimación no estandarizada. Los efectos indirecto ($a \times b$), directo (c') y total (c) se estimaron con intervalos percentiles bootstrap de 5000 réplicas, con reajuste del modelo latente en cada una; un efecto se considera distinto de cero cuando su intervalo excluye el cero. Se derivaron el R^2 de cada ecuación endógena y el f^2 de cada predictor. Cuarto, se evaluó la sensibilidad al estimador reestimando las trayectorias por mínimos cuadrados ordinarios sobre puntuaciones compuestas estandarizadas, con bootstrap de 5000 réplicas, y la estabilidad del efecto total ante el remuestreo (jackknife, submuestras del 80 % y del 60 %, y 1000 remuestreos con reemplazo).

4.4 RESULTADOS

4.4.1 Estadísticos descriptivos y correlaciones

Las tres puntuaciones compuestas mostraron medias cercanas al punto medio de la escala y distribuciones aproximadamente simétricas, con asimetrías entre –0.221 y 0.008 y curtosis entre –0.212 y 0.139 (Tabla 4-3; Figura 4-1A–B). Las correlaciones entre constructos fueron positivas y elevadas, con la mayor entre gobernanza y permanencia.

Tabla 4-3

Estadísticos descriptivos y correlaciones entre constructos

N.º	Variable	M	DE	Asimetría	Curtosis	1	2	3
1.	Gobernanza escolar	3.000	0.667	−0.011	−0.212	—	.792*** [.757, .823]	.736*** [.693, .774]
2.	Permanencia escolar	3.001	0.663	−0.221	0.139	.792***	—	.698*** [.651, .741]
3.	Rendimiento académico	3.000	0.667	0.008	−0.171	.736***	.698***	—

Nota: N = 500. Puntuaciones compuestas en la escala de respuesta de 1 a 5. Triángulo superior: correlación de Pearson con su intervalo de confianza al 95% (transformación Z de Fisher) entre corchetes; triángulo inferior: correlación de Pearson. ***p < .001.

4.4.2 Modelo de medida

Todas las cargas superaron .69, la fiabilidad fue adecuada (α entre .765 y .810; ω entre .765 y .809) y la AVE superó .50 en los tres constructos (Tabla 4-1). El valor máximo de la HTMT correspondió a gobernanza y permanencia y quedó por debajo del umbral estricto de .85 (Tabla 4-4; Figura 2A–B). El criterio de Fornell y Larcker se cumplió para el rendimiento académico, pero no para la gobernanza ni para la permanencia, cuyas raíces de AVE fueron inferiores a su correlación máxima.

Tabla 4-4

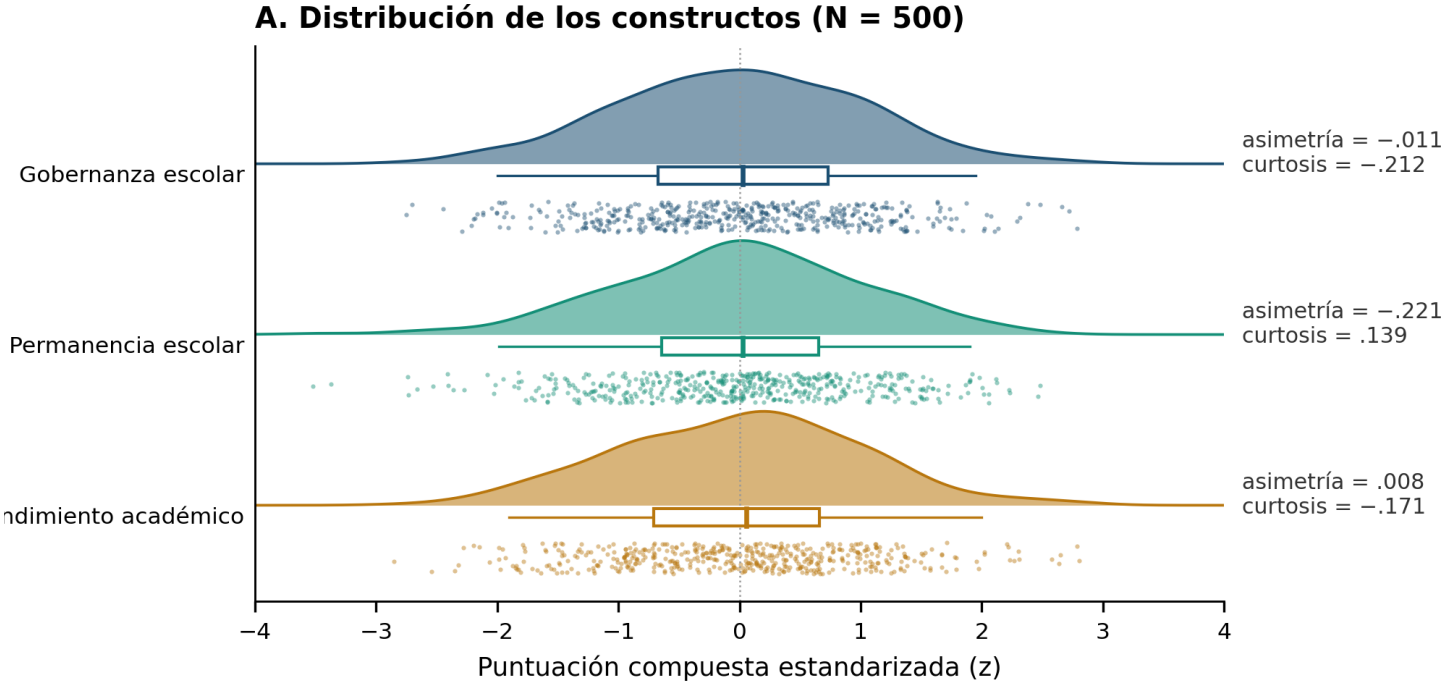
Validez discriminante: razón heterorrasgo-monorrasgo y criterio de Fornell-Larcker

Constructo	√AVE	Correlación máxima	Fornell-Larcker	HTMT con		
				gobernanza	permanencia	rendimiento
Gobernanza escolar	.764	.792	No se cumple	—	.833	.796
Permanencia escolar	.721	.792	No se cumple	.833	—	.727
Rendimiento académico	.766	.736	Se cumple	.796	.727	—

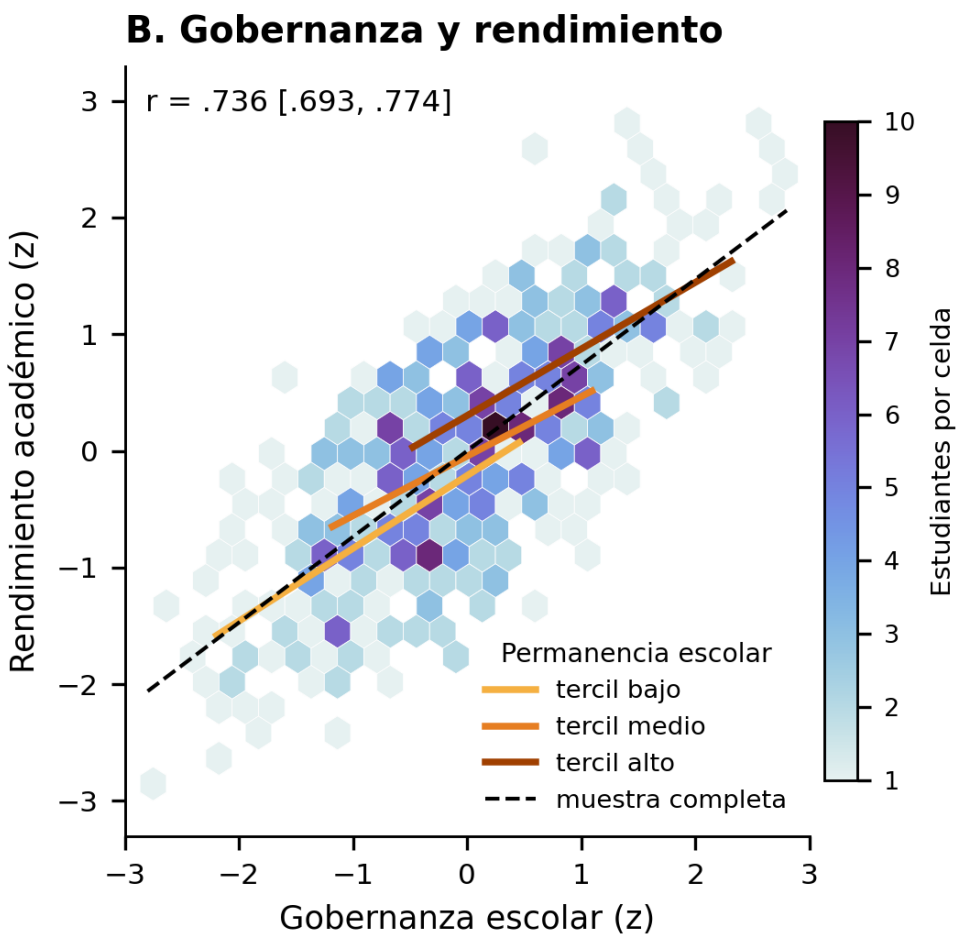
Nota: √AVE = raíz cuadrada de la varianza media extraída; correlación máxima = mayor correlación del constructo con los demás (Tabla 4-3). El criterio de Fornell y Larcker (1981) exige √AVE mayor que la correlación máxima. HTMT según Henseler et al. (2015): umbral estricto .85 y liberal .90; los tres pares cumplen ambos umbrales.

Figura 4-1

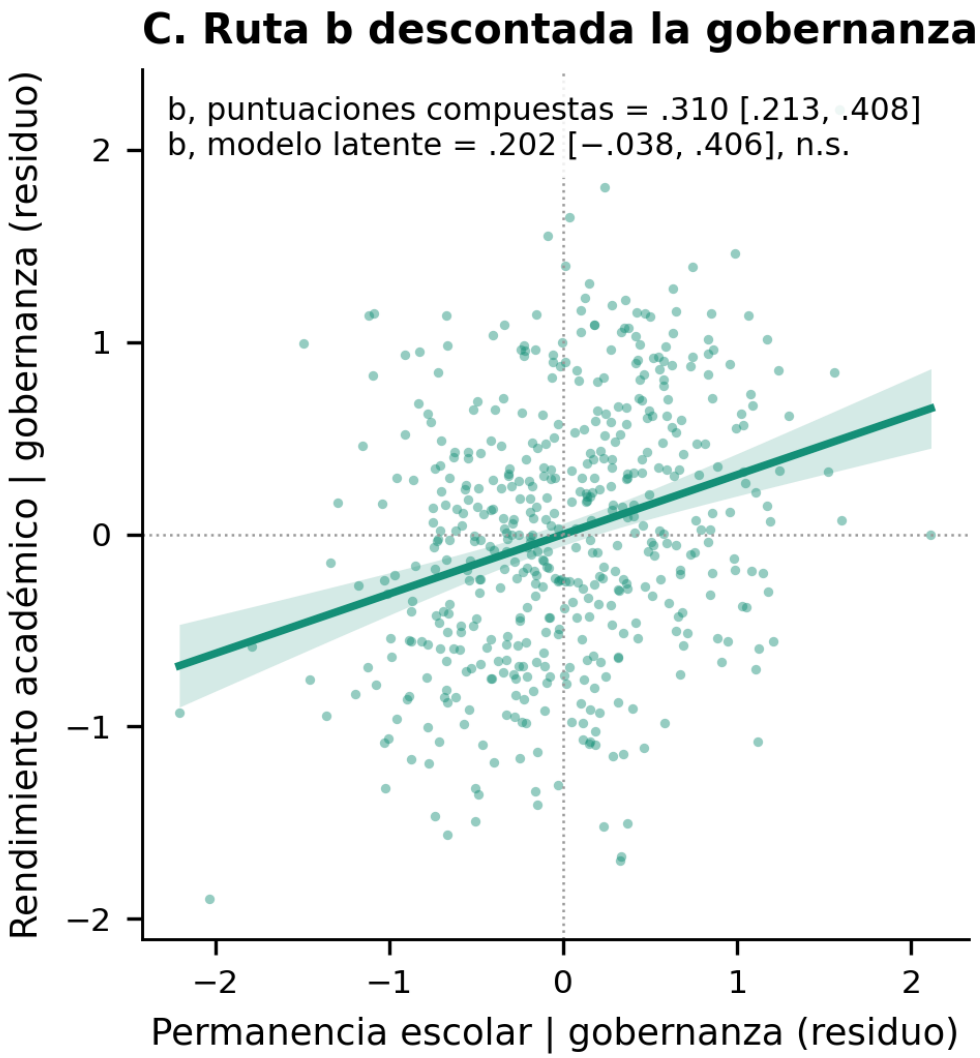
Distribución y relaciones observadas de los constructos



Nota. N = 500; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel A: densidad, caja (mediana y rango intercuartílico; bigotes del percentil 2.5 al 97.5) y observaciones individuales de cada constructo, con la asimetría y la curtosis de la Tabla 3 (gobernanza escolar: −.011 y −.212; permanencia escolar: −.221 y .139; rendimiento académico: .008 y −.171).



Nota. N = 500; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel B: densidad hexagonal de la gobernanza escolar y el rendimiento académico, con la recta de mínimos cuadrados de la muestra completa (discontinua) y las rectas descriptivas dentro de cada tercil de permanencia escolar; correlación de Pearson r = .736, IC95% [.693, .774].



Nota. N = 500; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel C: gráfico de variable añadida de la ruta b: residuos del rendimiento académico y de la permanencia escolar tras descontar la gobernanza escolar, con la recta de mínimos cuadrados y su banda de confianza al 95%; la pendiente es el coeficiente b de las puntuaciones compuestas, b = .310, IC95% [.213, .408], mientras que en el modelo latente b = .202, IC95% [−.038, .406], no significativo.

4.4.3 Ajuste del modelo estructural

El modelo hipotetizado presentó un ajuste adecuado según todos los índices y superó con claridad al modelo alternativo de un solo factor; la diferencia de χ^2 fue significativa, $\Delta\chi^2(3) = 160.482$, $p < .001$ (Tabla 4-5; Figura 4-2C). La Figura 4-2D representa la potencia de los criterios del cálculo a priori (Tabla 4-2) en función del tamaño de la muestra.

Tabla 4-5

Índices de ajuste del modelo hipotetizado y del modelo alternativo

Modelo	χ^2	gl	CFI	TLI	RMSEA	SRMR
Modelo hipotetizado (tres factores, mediación)	24.588	24	1.000	1.000	.007	.017
Modelo alternativo (un factor)	185.070	27	.920	.893	.108	.083
Diferencia (alternativo – hipotetizado)	$\Delta\chi^2 = 160.482$	$\Delta gl = 3$	−.080	−.107	+.101	+.066

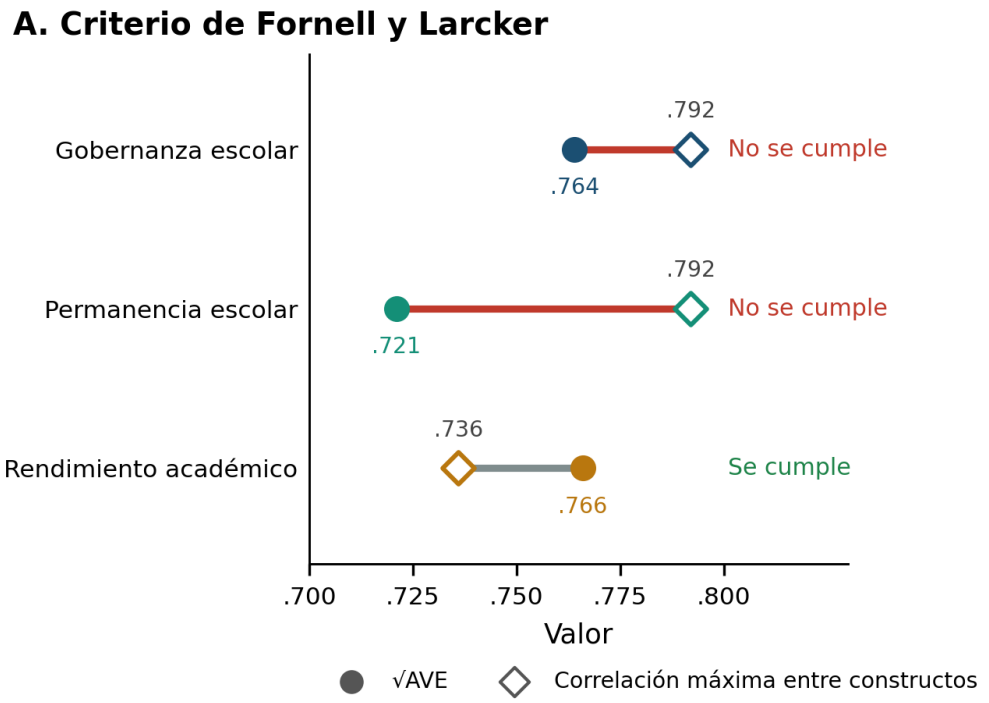
Nota: Estimación por máxima verosimilitud. Para el modelo hipotetizado, $p = .428$. La diferencia de χ^2 entre modelos anidados es significativa, $p < .001$. Puntos de corte de referencia: CFI y TLI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu & Bentler, 1999).

4.4.4 Modelo estructural y contraste de hipótesis

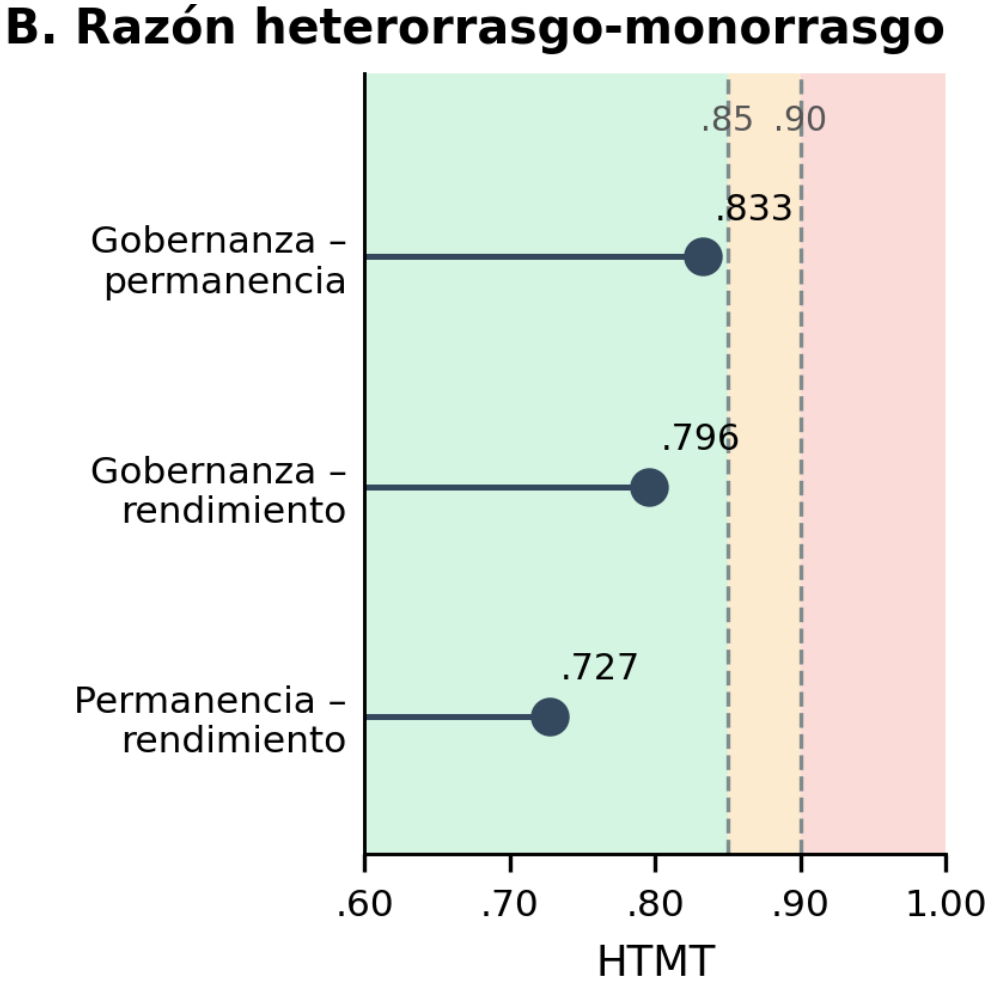
La gobernanza escolar se asoció positivamente con la permanencia (ruta a) y con el rendimiento académico (efecto directo c′), lo que apoya H1 (Tabla 4-6; Figura 4-3A). La ruta b, de la permanencia al rendimiento una vez considerada la gobernanza, fue positiva pero no alcanzó significación estadística. Como solo una de las dos rutas del mecanismo fue distinta de cero, H2 no se apoya.

Figura 4-2

Validez discriminante, ajuste comparado de los modelos y potencia del cálculo a priori del tamaño de la muestra

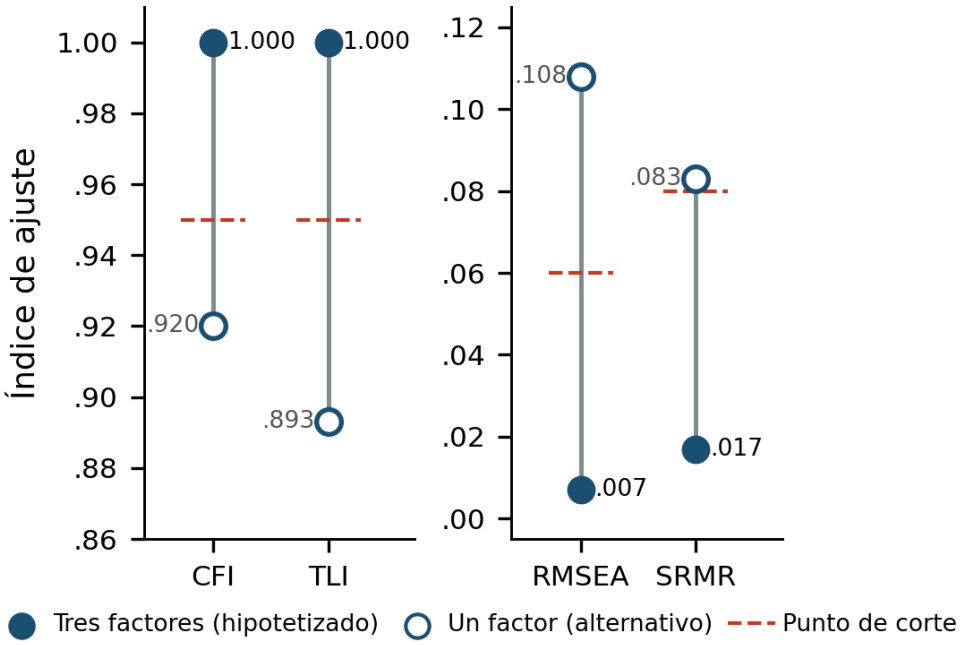


Nota. N = 500. Panel A: raíz cuadrada de la varianza media extraída (√AVE, círculo) y correlación máxima del constructo con los demás (rombo); el criterio de Fornell y Larcker no se cumple para la gobernanza escolar (.764 frente a .792) ni para la permanencia escolar (.721 frente a .792) y se cumple para el rendimiento académico (.766 frente a .736).



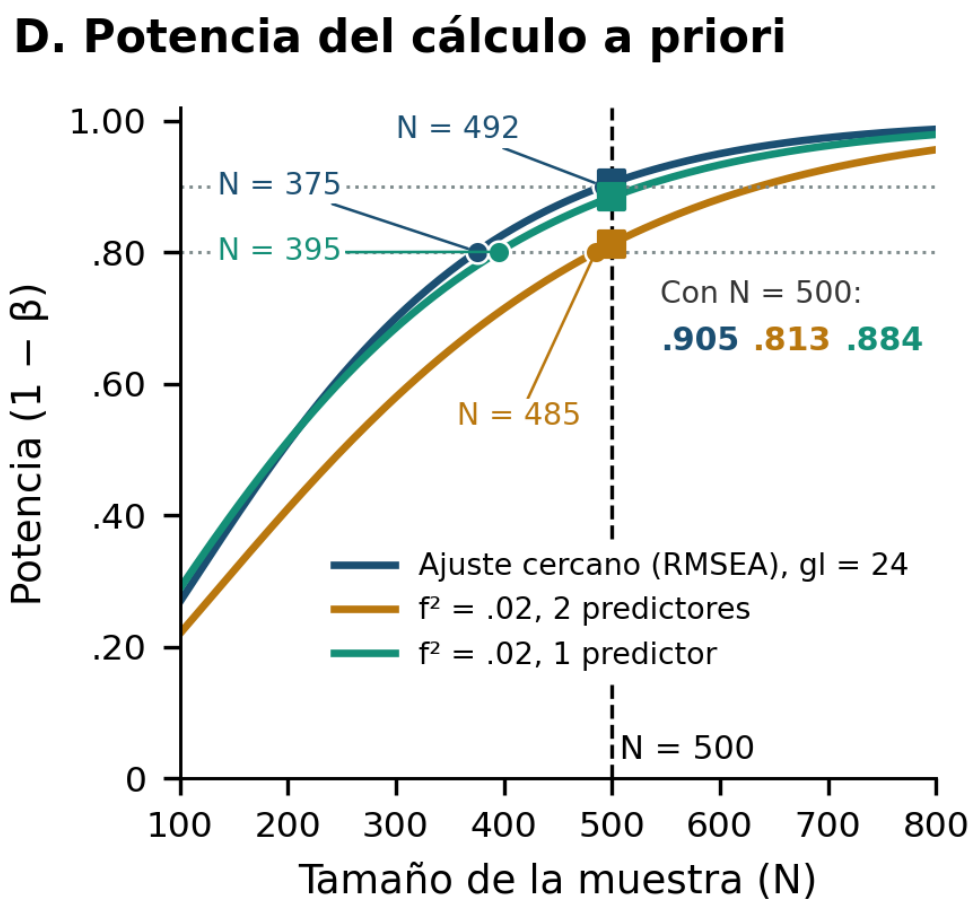
Nota. N = 500. Panel B: razón heterorrasgo-monorrasgo (HTMT) de cada par de constructos, con los umbrales .85 y .90: gobernanza–permanencia = .833, gobernanza–rendimiento = .796 y permanencia–rendimiento = .727.

C. Ajuste comparado de los modelos



$\Delta\chi^2(3) = 160.48$, $p < .001$ (un factor frente a tres factores)

Nota. Panel C: índices de ajuste del modelo hipotetizado de tres factores (círculo lleno: CFI = 1.000, TLI = 1.000, RMSEA = .007, SRMR = .017) frente al modelo alternativo de un factor (círculo vacío: CFI = .920, TLI = .893, RMSEA = .108, SRMR = .083), con los puntos de corte CFI y TLI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08; $\Delta\chi^2(3) = 160.48$, $p < .001$.



Nota. Panel D: potencia en función del tamaño de la muestra para la prueba de ajuste cercano del RMSEA (gl = 24; ε0 = .05; ε1 = .08; α = .05) y para un efecto f^2 = .02 con dos predictores y con un predictor (α = .05); los círculos marcan el N mínimo de cada criterio (492 para una potencia de .90 y 375 para .80 en el RMSEA; 485 y 395 para f^2 = .02 con dos predictores y con un predictor) y los cuadrados la potencia con N = 500 (.905, .813 y .884).

Tabla 4-6

Coeficientes estructurales del modelo

Ruta	β	EE	p	Hipótesis	Decisión
Gobernanza escolar → Rendimiento académico (c′)	.605	.111	< .001	H1	Apoyada
Gobernanza escolar → Permanencia escolar (a)	.824	.064	< .001	H2 (ruta a)	Ruta significativa
Permanencia escolar → Rendimiento académico (b)	.202	.116	.085	H2 (ruta b)	Ruta no significativa; H2 no apoyada
Gobernanza escolar → Permanencia escolar → Rendimiento académico (a × b)	.166	—	.088	H3	No apoyada

Nota: β = coeficiente estandarizado; EE y p de la estimación no estandarizada por máxima verosimilitud. La decisión sobre H2 se basa en la significación conjunta de las rutas a y b; la decisión sobre H3, en el intervalo bootstrap del efecto indirecto (Tabla 4-7).

La descomposición de efectos (Tabla 4-7; Figuras 4-3A y 4-3C) mostró un efecto total positivo, dominado por el componente directo. El efecto indirecto fue positivo, pero su intervalo bootstrap incluyó el cero; por tanto, H3 no se apoya con el estimador latente. El efecto indirecto representó el 21.5 % del efecto total.

Tabla 4-7

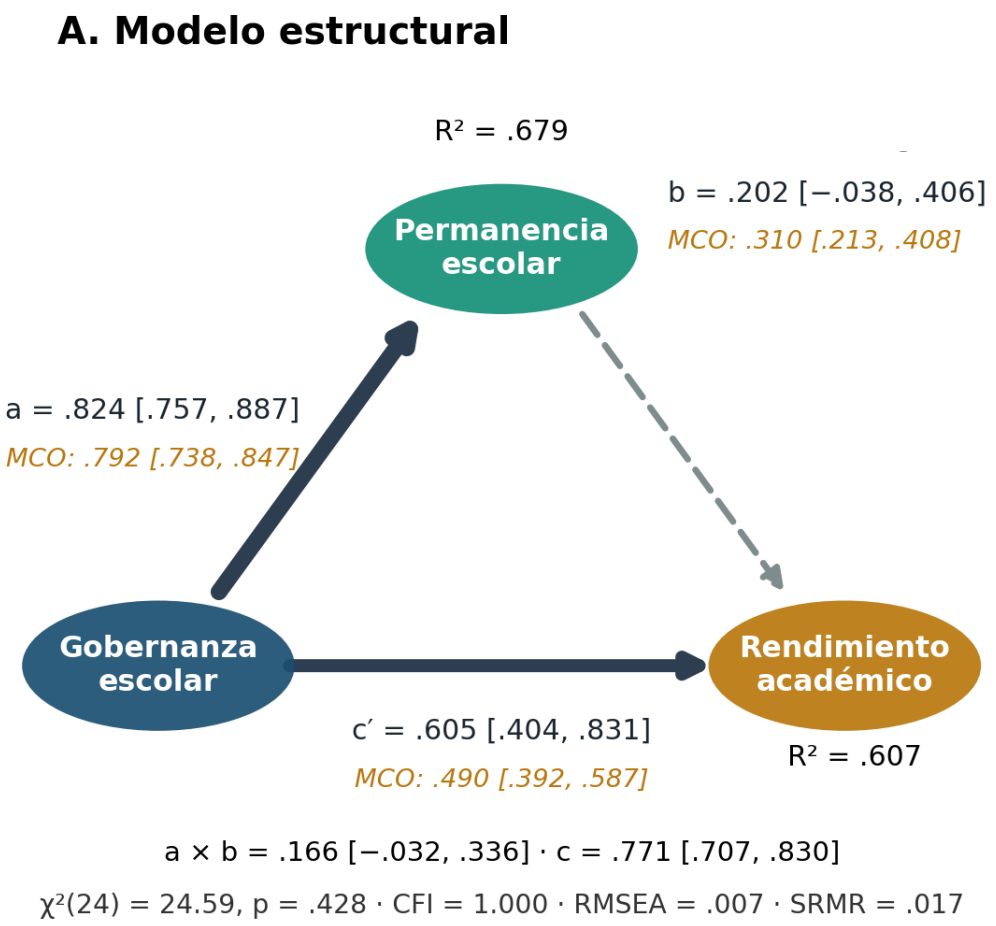
Descomposición del efecto de la gobernanza escolar sobre el rendimiento académico

Efecto		β	IC 95 %		Distinto de cero
			inferior	superior	
a:	gobernanza → permanencia	.824	.757	.887	Sí
b:	permanencia → rendimiento	.202	−.038	.406	No
c′:	efecto directo	.605	.404	.831	Sí
a × b:	efecto indirecto	.166	−.032	.336	No
c:	efecto total (c′ + a × b)	.771	.707	.830	Sí

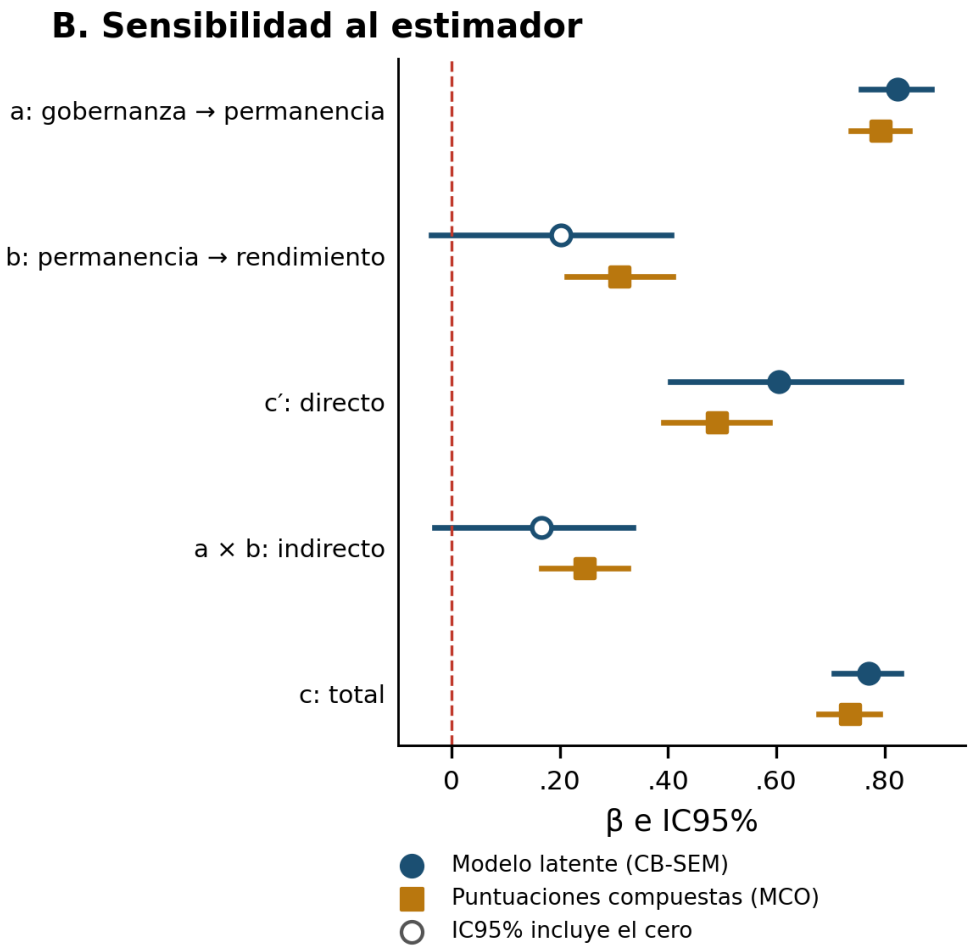
Nota: Intervalos de confianza percentiles bootstrap de 5000 réplicas (5000 réplicas convergentes), con reajuste del modelo latente en cada réplica. Un efecto se considera distinto de cero cuando su intervalo excluye el cero.

Figura 4-3

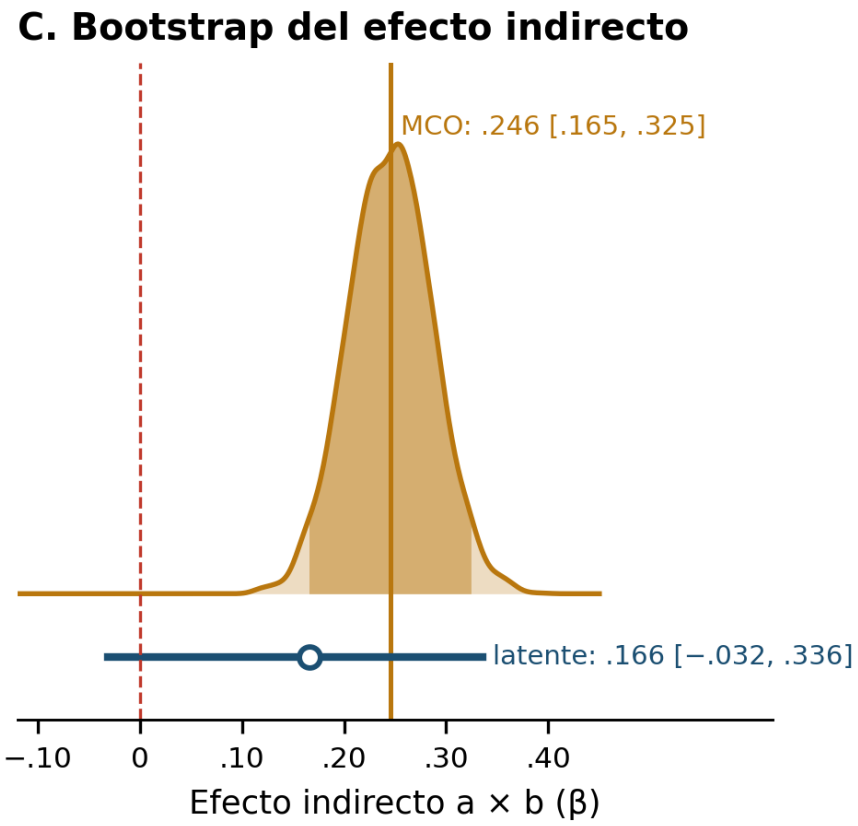
Modelo estructural, descomposición del efecto y sensibilidad al estimador



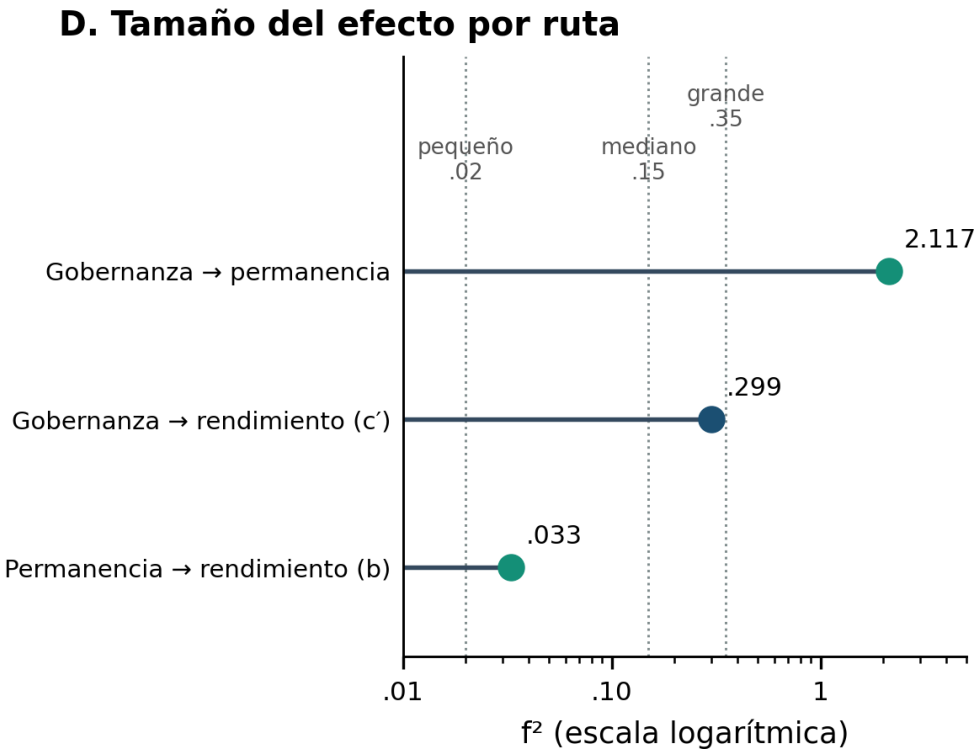
Nota. N = 500. Panel A: modelo de ecuaciones estructurales basado en covarianzas, estimado por máxima verosimilitud; coeficientes estandarizados con intervalos de confianza bootstrap percentiles de 5000 réplicas y, en cursiva, la reestimación por mínimos cuadrados ordinarios (MCO) sobre puntuaciones compuestas. El grosor de cada flecha es proporcional al coeficiente del modelo latente y la flecha discontinua indica una ruta no significativa. Ruta a: β = .824 [.757, .887]; ruta b: β = .202 [−.038, .406]; efecto directo c′: β = .605 [.404, .831]; efecto indirecto a × b = .166 [−.032, .336]; efecto total c = .771 [.707, .830]; R^2 = .679 para la permanencia escolar y .607 para el rendimiento académico. Ajuste: $\chi^2(24)$ = 24.59, p = .428, CFI = 1.000, RMSEA = .007, SRMR = .017.



Nota. N = 500. Panel B: cada efecto con los dos estimadores (círculo: modelo latente; cuadrado: puntuaciones compuestas) y su IC95%; los marcadores vacíos señalan intervalos que incluyen el cero (ruta b y efecto indirecto del modelo latente).



Nota. Panel C: distribución de las 5000 réplicas bootstrap del efecto indirecto con puntuaciones compuestas (β = .246; intervalo percentil sombreado [.165, .325]) frente al intervalo del modelo latente (β = .166, IC95% [−.032, .336]), que incluye el cero.



Nota. Panel D: tamaño del efecto f^2 de cada ruta en escala logarítmica, con los umbrales .02, .15 y .35: gobernanza → permanencia, $f^2 = 2.117$; gobernanza → rendimiento (c′), $f^2 = .299$; permanencia → rendimiento (b), $f^2 = .033$.

4.4.5 Varianza explicada y tamaños del efecto

El modelo explicó el 67.9 % de la varianza de la permanencia y el 60.7 % de la del rendimiento (Tabla 4-8; Figura 4-3D). En la ecuación del rendimiento, el aporte exclusivo de la permanencia fue pequeño frente al de la gobernanza.

Tabla 4-8

Varianza explicada y tamaño del efecto por ruta

Ecuación	Predictor	β	R^2		f^2
			de la ecuación	sin el predictor	
Permanencia escolar	Gobernanza escolar	.824	.679	.000	2.117
Rendimiento académico	Gobernanza escolar (directo)	.605	.607	.490	.299
Rendimiento académico	Permanencia escolar	.202	.607	.594	.033

Nota: R^2 derivados de los coeficientes estandarizados del modelo estructural: $R^2(\text{permanencia}) = a^2$; $R^2(\text{rendimiento}) = c'^2 + b^2 + 2 \cdot c' \cdot b \cdot a$. $f^2 = (R^2 \text{ con el predictor} - R^2 \text{ sin el predictor}) / (1 - R^2 \text{ con el predictor})$.

4.4.6 Sensibilidad al estimador y estabilidad

La reestimación por mínimos cuadrados ordinarios sobre puntuaciones compuestas reprodujo la dirección de todas las rutas (Tabla 4-9; Figuras 4-1C y 4-3B). En este estimador, que no corrige el error de medida, la ruta b fue significativa y el efecto indirecto resultó distinto de cero. Las diferencias con el modelo latente se concentraron en el segundo tramo del mecanismo: con el modelo latente aumentó la correlación entre gobernanza y permanencia y se amplió la incertidumbre de la ruta b.

Tabla 4-9

Sensibilidad de los efectos al estimador

Efecto		Modelo latente:		Puntuaciones compuestas:	
		β	IC 95 %	β	IC 95 %
a:	gobernanza → permanencia	.824	[.757, .887]	.792	[.738, .847]
b:	permanencia → rendimiento	.202	[−.038, .406]	.310	[.213, .408]
c′:	efecto directo	.605	[.404, .831]	.490	[.392, .587]
a × b:	efecto indirecto	.166	[−.032, .336]	.246	[.165, .325]
c:	efecto total	.771	[.707, .830]	.736	[.679, .791]

Nota: Modelo latente: ecuaciones estructurales por máxima verosimilitud. Puntuaciones compuestas: mínimos cuadrados ordinarios sobre la media de los ítems de cada escala, estandarizada. En ambos estimadores, intervalos percentiles bootstrap de 5000 réplicas.

El efecto total en su métrica original (b = 0.466) se mantuvo estable en todas las perturbaciones de la muestra: signo y significación se conservaron en el 100 % de las réplicas, con desviaciones estándar entre 0.0008 y 0.0184 (Tabla 4-10).

Tabla 4-10

Estabilidad del efecto total ante el remuestreo de la muestra

Procedimiento	Réplicas	b medio	DE	IC 95 %	Signo estable (%)	Significativo (%)
Jackknife (eliminación de un caso)	500	0.466	0.0008	[0.464, 0.468]	100 %	100 %
Submuestras del 80%	100	0.465	0.0090	[0.449, 0.483]	100 %	100 %
Submuestras del 60%	100	0.466	0.0159	[0.432, 0.492]	100 %	100 %
Remuestreo con reemplazo	1000	0.466	0.0184	[0.432, 0.503]	100 %	100 %

Nota: b = coeficiente no estandarizado del efecto total de la gobernanza escolar sobre el rendimiento académico en la regresión por mínimos cuadrados ordinarios sobre puntuaciones compuestas (b = 0.466, p < .001). Significativo: porcentaje de réplicas con p < .05.

4.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre gobernanza y rendimiento, sino que el mecanismo de retención que los sistemas de gestión por compromisos suponen se desvanece o se confirma según el estimador. La primera ruta fue intensa y la segunda, débil; el efecto indirecto fue positivo en ambos estimadores, pero solo se distinguió de cero con puntuaciones compuestas, que no corrigen el error de medida. Así, los resultados apoyan H1 y no apoyan H2 ni H3 con el estimador latente, y esa inestabilidad condiciona la lectura de todo el modelo: la permanencia es un mecanismo plausible, pero no concluyente.

Conviene examinar primero la magnitud del vínculo directo. Las síntesis sobre liderazgo directivo y clima escolar sitúan la asociación con el logro en efectos de 0.08 a 0.16 desviaciones estándar (Liebowitz & Porter, 2019), en correlaciones pequeñas (Demirtas et al., 2021), en tamaños de efecto pequeños para el clima de aula y medios para el clima escolar (Erdem & Kaya, 2023) y en estimaciones meta-metaanalíticas de magnitud modesta (Wu & Shen, 2022); un modelo estructural con cinco factores de clima explicó apenas el 19 % de la variación del desempeño (Amsalu & Belay, 2024). El efecto total aquí observado supera con holgura ese rango. Una explicación razonable, aunque no contrastada en este diseño, es la varianza compartida de método: gobernanza y rendimiento proceden del mismo informante, y las percepciones del clima se relacionan con otras percepciones psicosociales del estudiante, como la esperanza y la motivación académica (Gentzis & Dixson, 2023). La naturaleza del informante también importa, pues la comparación de perspectivas de directivos y docentes sobre liderazgo y logro revela lecturas diferenciadas del mismo fenómeno (Zhang & Wu, 2025). Es consistente con ello que la magnitud deba leerse como la intensidad con que los estudiantes vinculan las condiciones institucionales con su propio desempeño, y no como una estimación del efecto de la política sobre aprendizajes medidos externamente. A ello se suma que la estabilidad del efecto total ante el

remuestreo descarta que el resultado dependa de unos pocos casos, pero no que dependa de la forma de medir.

Menos esperada fue la debilidad de la ruta b. La literatura longitudinal sobre pertenencia sugiere que el sentido de pertenencia antecede al rendimiento, con efectos recíprocos que parecen mayores en las niñas (Sakellariou, 2025), y la intención de abandono se asocia con factores socio-psicológicos, entre ellos las dificultades de atención que deterioran los hábitos de estudio, que la escuela puede modificar (Ruzzante et al., 2026). Desde ese marco cabría esperar una contribución más nítida de la permanencia. Sin embargo, nuestra medida captura la intención de continuar en estudiantes que ya permanecen matriculados, una población con poca variabilidad en el extremo de abandono; en ellos, la intención de continuar puede funcionar más como condición necesaria que como impulsor del desempeño. Hajovsky y Chesnut (2025) hallaron, además, que la variabilidad del logro no predijo cambios posteriores en la relación docente–estudiante, lo que sugiere asimetrías temporales que un corte transversal no puede resolver: si la relación entre permanencia y logro es recíproca, la ruta b estimada en un solo momento promedia influjos de signo y retardo distintos. Cabe, por último, que la ruta b esté parcialmente absorbida por la gobernanza, dada la proporción elevada de varianza que ambos constructos comparten y el incumplimiento del criterio de Fornell y Larcker; la separación entre ambos descansa en la HTMT, y toda estimación de b condicionada a la gobernanza hereda esa fragilidad.

Si la retención no es el canal dominante, la gobernanza debería operar sobre el rendimiento por vías directas, y la literatura sobre liderazgo instruccional apunta en esa dirección. Promover el aprendizaje docente tuvo una influencia dominante sobre la calidad de la enseñanza en escuelas rurales chinas, mientras que gestionar la instrucción tuvo un efecto significativo pero débil (Lin & Hu, 2025); el liderazgo pedagógico del director incide en el compromiso docente a través del clima cultural de la escuela (Li et al., 2026); y un programa de renovación del liderazgo escolar elevó el crecimiento anual de la proficiencia en lengua en las escuelas experimentales (Ma et al., 2024). Estos canales, centrados en el

tiempo efectivo de enseñanza y el seguimiento del aprendizaje, coinciden con las condiciones de funcionamiento que miden los compromisos peruanos, aunque nuestro diseño no permite aislar cuál de ellas explica el componente directo.

La sensibilidad del efecto indirecto al estimador tiene un alcance metodológico que trasciende el caso. Al corregir el error de medida, la correlación entre gobernanza y permanencia aumenta y la ruta b se estima con mayor incertidumbre; con puntuaciones compuestas, la atenuación reduce esa dependencia y el efecto indirecto se vuelve significativo. Las evaluaciones de mediación en política educativa que se basan en un único estimador pueden, por tanto, sobrestimar o subestimar la evidencia sobre el canal de retención. Esto es coherente con los hallazgos multinivel recientes según los cuales los mecanismos del liderazgo varían en magnitud y en vía según el contexto escolar (Wu et al., 2026).

En la misma línea de los modelos estructurales sobre clima institucional, el resultado respecto del efecto total concuerda con la evidencia que identifica el liderazgo y la disciplina escolar entre los correlatos del logro (Lee et al., 2026), lo que refuerza la plausibilidad del vínculo directo.

Para la gestión escolar en el Perú y en sistemas con arquitecturas similares, los resultados sugieren que los compromisos de condiciones de funcionamiento —calendario, acompañamiento y convivencia— se asocian con el rendimiento con independencia de su asociación con la permanencia. La evaluación de estos compromisos no debería, entonces, subordinar su valor a la reducción del abandono, sino examinar por separado su relación con el aprendizaje y con la trayectoria del estudiante. La revisión de la gobernanza de nivel intermedio en América Latina advierte que pocos estudios evalúan si y cómo estas estructuras mejoran la enseñanza y el aprendizaje, de modo que la medición permanece poco desarrollada (Asim et al., 2026), y la evidencia chilena sobre eficacia escolar sugiere que las escuelas de baja efectividad deben fortalecer su capacidad de promover el aprendizaje (Kuzmanic et al., 2024); ambos argumentos

respaldan la necesidad de indicadores institucionales que acompañen a las percepciones estudiantiles.

El estudio presenta limitaciones que acotan sus conclusiones. El diseño transversal no permite establecer el orden temporal, de modo que la mediación se interpreta como un patrón de asociación compatible con el mecanismo propuesto y no como una secuencia causal. Las tres escalas son autoinformes, por lo que la varianza común de método puede haber inflado las asociaciones, y el rendimiento autopercibido no sustituye a las calificaciones ni a las pruebas estandarizadas. La muestra se obtuvo por conveniencia, lo que impide generalizar al conjunto de estudiantes de secundaria pública. El análisis se realizó en un solo nivel, sin modelar el anidamiento de los estudiantes en escuelas, de modo que la gobernanza se interpreta como percepción individual y no como atributo de la institución. Por último, gobernanza y permanencia comparten una proporción elevada de varianza, y su validez discriminante se sostiene con la HTMT, pero no con el criterio de Fornell y Larcker. Estudios futuros deberían combinar indicadores institucionales de cumplimiento de los compromisos, registros administrativos de matrícula y resultados de evaluaciones estandarizadas en diseños longitudinales y multinivel.

4.6 CONCLUSIONES

La gobernanza institucional organizada en torno a los Compromisos de Gestión Escolar se asocia de forma intensa y estable con el rendimiento académico de los estudiantes de secundaria pública del Perú, con un efecto total de $\beta = .771$ y un componente directo de $\beta = .605$. La gobernanza también se asocia con la intención de permanencia ($\beta = .824$), pero la permanencia transmite solo una fracción de esa asociación con el rendimiento (β indirecto $= .166$), y dicha fracción no es distinta de cero cuando se corrige el error de medida. En términos de hipótesis, el efecto directo se sostiene, mientras que ni la significación conjunta de las dos rutas ni el intervalo bootstrap del efecto indirecto respaldan la mediación con el estimador latente.

La contribución del estudio consiste en mostrar que el canal de retención implícito en los sistemas de gestión por compromisos es plausible pero no concluyente, y que su detección depende del estimador empleado. Este resultado tiene una implicación práctica: las políticas de gestión escolar deben evaluarse por su relación directa con el aprendizaje y no solo por su capacidad de retener estudiantes, y los análisis de mediación en este campo deberían reportar la sensibilidad de sus conclusiones al tratamiento del error de medida.

Las conclusiones se circunscriben a una muestra por conveniencia, a medidas de autoinforme y a un diseño transversal de un solo nivel. La investigación posterior debería contrastar el papel de la permanencia como mecanismo con diseños longitudinales, modelos multinivel que incorporen la escuela como unidad de análisis y medidas externas del logro, de modo que pueda establecerse si la retención antecede al aprendizaje, lo sucede o ambos se refuerzan mutuamente.

4.7 REFERENCIAS BIBLIOGRÁFICAS

Adrogué, C., & Orlicki, E. (2025). Riesgo de abandono escolar en la secundaria. Factores comunes y diferentes para varones y mujeres. *Estudios económicos, 42*(85), 58-79. https://doi.org/10.52292/j.estudecon.2025.5069

Amsalu, A., & Belay, S. (2024). Analyzing the Contribution of School Climate to Academic Achievement Using Structural Equation Modeling. *Sage Open, 14*(1). https://doi.org/10.1177/21582440241227271

Asim, M., Dalmon, D., & Núñez, I. (2026). What Do We Know About the Middle Tier of Education Governance from Latin America? *Review of Educational Research*. https://doi.org/10.3102/00346543261462174

Barboza, A., Huamani, J., Santisteban, J., Maculan, N., & Mauricio, D. (2026). Prediction and Explainability of School Dropout in Peru

Using Machine Learning and Explainable Artificial Intelligence. *IEEE Access*, *14*, 43752-43771. https://doi.org/10.1109/ACCESS.2026.3675564

Borchers, J., & Cunha, M. (2025). School performance and inequality of opportunities in Latin America. *Studies in Educational Evaluation*, *86*, 101497. https://doi.org/10.1016/j.stueduc.2025.101497

Cabra, H. (2023). Three approaches to modeling the relationship among durable goods, academic achievement, and school attendance in Colombia1. *Heliyon*, *9*(12), e22732. https://doi.org/10.1016/j.heliyon.2023.e22732

Calero, R., Sosa, M., Lino, L., & Ponciano, J. (2023). Factores determinantes de la deserción escolar en la región Huánuco, Perú. *Desafíos*, *14*(2), 118-123. https://doi.org/10.37711/desafios.2023.14.2.401

Chen, D., & Bos, W. (2023). Relationship between principal leadership and student achievement: A comparative study between Germany and China. *International Journal of Educational Development*, *102*, 102868. https://doi.org/10.1016/j.ijedudev.2023.102868

Chocce, F. (2026). Desempeño docente y logros de aprendizaje en estudiantes de secundaria rural del Perú. *Llimpi*, *6*(1), 51-56. https://doi.org/10.54943/lree.v6i1.811

Demirtas, S., Akin, C., & Terzi, R. (2021). Does school climate that includes students' views deliver academic achievement? A multilevel meta-analysis. *School Effectiveness and School Improvement, 32*(4), 543-563. https://doi.org/10.1080/09243453.2021.1920432

Erdem, C., & Kaya, M. (2023). The relationship between school and classroom climate, and academic achievement: A meta-analysis. *School Psychology International*, *45*(4), 380-408. https://doi.org/10.1177/01430343231202923

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Gentzis, E., & Dixson, D. (2023). How School Climate Relates to Other Psychosocial Perceptions and Academic Achievement across the School Year. *School Psychology Review*, *53*(5), 441-458. https://doi.org/10.1080/2372966X.2023.2261835

Gómez, L., & Coronel, J. (2023). Compromisos de gestión escolar y administración de recursos en directivos de Educación Básica, Luya - 2022. *Revista Científica UNTRM: Ciencias Sociales y Humanidades*, 22-32. https://doi.org/10.25127/rcsh.20236.930

Hajovsky, D., & Chesnut, S. (2025). A longitudinal examination of parallel growth and reciprocal changes in teacher–student relationships and academic achievement. *School Psychology*, *40*(1), 1-12. https://doi.org/10.1037/spq0000583

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, L., & Wang, D. (2023). Teacher Support, Academic Self-Efficacy, Student Engagement, and Academic Achievement in Emergency Online Learning. *Behavioral Sciences, 13*(9), 704. https://doi.org/10.3390/bs13090704

Kuzmanic, D., Meneses, F., Valenzuela, J., Rodríguez, P., & Claro, S. (2024). Inequalities in learning loss during the COVID-19 pandemic in Chile: the significance of school effectiveness. *School Effectiveness and School Improvement, 35*(2), 95-115. https://doi.org/10.1080/09243453.2024.2328016

Lee, H., Lee, C., & Cham, H. (2026). Associations Between School Climate and Student Mathematics Achievement: A Multilevel Structural Equation Modeling Approach. *Education Sciences, 16*(2), 293. https://doi.org/10.3390/educsci16020293

Li, X., Gong, Y., Zhou, J., & Zhao, P. (2026). Does Principal Instructional Leadership Promote Teacher Instructional Engagement? Evidence from a Large-Scale Survey in China. *Education Sciences, 16*(3), 429. https://doi.org/10.3390/educsci16030429

Liebowitz, D., & Porter, L. (2019). The Effect of Principal Behaviors on Student, Teacher, and School Outcomes: A Systematic Review and Meta-Analysis of the Empirical Literature. *Review of Educational Research, 89*(5), 785-827. https://doi.org/10.3102/0034654319866133

Lin, Y., & Hu, D. (2025). Instructional leadership malfunction in rural areas: Examining the influence of principal instructional leadership on instructional quality in Chinese rural schools. *Educational Management Administration & Leadership, 54*(5), 1572-1589. https://doi.org/10.1177/17411432251346085

López, P. (2026). Public educational policies, school entrepreneurship, and student motivation in public schools of metropolitan Lima-Peru. *Multidisciplinary Science Journal, 9*(3), 2027114. https://doi.org/10.31893/multiscience.2027114

Ma, X., Shen, J., Reeves, P., Wu, H., Roberts, L., Zheng, Y., & Chen, Q. (2024). Effects of the "High Impact Leadership for School Renewal" Project on Principal Leadership, School Leadership, and Student Achievement. *Education Sciences, 14*(6), 600. https://doi.org/10.3390/educsci14060600

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Magallanes, F., Aguirre, C., & Magallanes, M. (2024). La Gestión Educativa y la Convivencia Escolar en las instituciones educativas del Perú. *Igobernanza, 7*(27), 246-263. https://doi.org/10.47865/igob.vol7.n27.2024.362

Mendoza, M., & Ballesta, E. (2023). Más allá de la deserción escolar: Importancia atribuida a la finalización de la educación secundaria en cuatro comunidades educativas chilenas. *Actualidades Investigativas en Educación, 23*(3), 1-30. https://doi.org/10.15517/aie.v23i3.54644

Ruzzante, F., Loconte, R., Romeo, M., Sesso, G., D'Arcangelo, S., Maggi, F., Ricciardi, E., & Pietrini, P. (2026). Socio-psychological predictors of school dropout intention among Italian adolescents: evidence from a large-scale study. *Frontiers in Psychology*, *17*. https://doi.org/10.3389/fpsyg.2026.1705896

Sakellariou, C. (2025). Reciprocal longitudinal effects between sense of school belonging and academic achievement: quasi-experimental estimates using United States primary school data. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1478320

Solís, A., Berrocal, M., & Guzmán, S. (2024). Estrategias de gestión institucional para fortalecer la calidad educativa en escuelas públicas del sur peruano. *REVISTA EDUSER*, *11*(2), 58-67. https://doi.org/10.18050/eduser.v11n2a6

Song, S., Martin, M., & Wang, Z. (2024). School belonging mediates the longitudinal effects of racial/ethnic identity on academic achievement and emotional well-being among Black and Latinx adolescents. *Journal of School Psychology*, *106*, 101330. https://doi.org/10.1016/j.jsp.2024.101330

Székely, M., & Karver, J. (2021). Youth Out of School and Out of Work in Latin America: A Cohort Approach. *International Journal of Educational Development*, *80*, 102294. https://doi.org/10.1016/j.ijedudev.2020.102294

Tao, Y., Meng, Y., Gao, Z., & Yang, X. (2022). Perceived teacher support, student engagement, and academic achievement: a meta-analysis. *Educational Psychology*, *42*(4), 401-420. https://doi.org/10.1080/01443410.2022.2033168

Wu, H., & Shen, J. (2022). The association between principal leadership and student achievement: A multivariate meta-meta-analysis. *Educational Research Review*, *35*, 100423. https://doi.org/10.1016/j.edurev.2021.100423

Wu, H., & Shen, J. (2025). Validating the Learning-Centered School Leadership Framework for Predicting Student Achievement: Evidence From Meta-Analysis. *Journal of School Leadership*, *36*(1), 116-136. https://doi.org/10.1177/10526846251374239

Wu, H., Zhang, S., & Marszalek, J. (2026). Principal leadership, school processes, and student achievement: Examining contextual heterogeneity in leadership mechanisms via multigroup multilevel structural equation modeling. *Educational Management Administration & Leadership*. https://doi.org/10.1177/17411432261434762

Zhang, S., & Wu, H. (2025). School Leadership, Parental Involvement, and Student Achievement: A Comparative Analysis of Principal and Teacher Perspectives. *Education Sciences*, *15*(6), 767. https://doi.org/10.3390/educsci15060767

Zhou, J., Gong, X., & Li, X. (2024). Longitudinal relations between teacher support and academic achievement among Chinese children: Disentangling between- and within-student associations. *Journal of School Psychology*, *103*, 101287. https://doi.org/10.1016/j.jsp.2024.101287

CAPÍTULO V

CONOCIMIENTO DEL RAZONAMIENTO ALGEBRAICO TEMPRANO, AUTOEFICACIA DOCENTE E INTEGRACIÓN DEL PENSAMIENTO ALGEBRAICO EN LA ENSEÑANZA DE LAS CIENCIAS: UN MODELO DE ECUACIONES ESTRUCTURALES EN LA FORMACIÓN INICIAL DEL PROFESORADO

Romer Juvenal Javier Quijano

Universidad Nacional Hermilio Valdizán

Resumen

La formación inicial del profesorado de ciencias suele desarticular el conocimiento algebraico de su uso en el aula. Este estudio examinó en qué medida el conocimiento del razonamiento algebraico temprano se asocia con la integración del pensamiento algebraico en la enseñanza de las ciencias y qué parte de esa asociación se transmite mediante la autoeficacia docente. Con un diseño no experimental, transversal y explicativo, se analizó a 396 estudiantes peruanos de formación inicial docente en ciencias, seleccionados por conveniencia durante 2025. Se estimó un modelo de ecuaciones estructurales basado en covarianzas, con tres constructos medidos mediante escalas de tres ítems, intervalos bootstrap de 5000 réplicas y análisis de robustez por regresión jerárquica y pruebas de estrés. El modelo de medición mostró cargas de .711 a .876 y validez convergente y discriminante, y el modelo estructural ajustó adecuadamente (CFI = .997, RMSEA = .024, SRMR = .022). El conocimiento se asoció con la autoeficacia (β = .660), la autoeficacia con la integración (β = .317) y el conocimiento conservó un efecto directo (β = .296). El efecto indirecto fue de .209 (IC 95 % [.095, .335]), el 41.4 % del efecto total, lo que configura una mediación parcial. Los resultados sugieren que los programas de formación deberían intervenir de forma simultánea sobre el conocimiento algebraico y sobre la autoeficacia para enseñarlo, con la cautela que impone un diseño transversal basado en autoinforme.

Palabras clave: razonamiento algebraico temprano; autoeficacia docente; formación inicial docente; enseñanza de las ciencias; ecuaciones estructurales; mediación

Abstract

Initial science teacher education often detaches algebraic knowledge from its classroom use. This study examined to what extent knowledge of early algebraic reasoning is associated with the integration of algebraic thinking into science

teaching, and what part of that association is transmitted through teacher self-efficacy. Using a non-experimental, cross-sectional, explanatory design, we analyzed 396 Peruvian pre-service science teachers, selected by convenience sampling in 2025. A covariance-based structural equation model was estimated with three constructs measured by three-item scales, bootstrap intervals of 5000 resamples, and robustness analyses through hierarchical regression and stress tests. The measurement model showed loadings between .711 and .876 and convergent and discriminant validity, and the structural model fit adequately (CFI = .997, RMSEA = .024, SRMR = .022). Knowledge was associated with self-efficacy (β = .660), self-efficacy with integration (β = .317), and knowledge retained a direct effect (β = .296). The indirect effect was .209 (95% CI [.095, .335]), 41.4% of the total effect, which indicates partial mediation. The results suggest that teacher education programs should intervene simultaneously on algebraic knowledge and on self-efficacy for teaching it, with the caution imposed by a cross-sectional, self-report design.

Keywords: *early algebraic reasoning; teacher self-efficacy; initial teacher education; science teaching; structural equation modeling; mediation*

5.1 INTRODUCCIÓN

La formación inicial del profesorado de ciencias enfrenta una tensión persistente entre el dominio de los contenidos científicos y la capacidad de movilizar el razonamiento algebraico temprano —la generalización de patrones, el uso de relaciones funcionales entre magnitudes y su expresión simbólica— como herramienta para modelar fenómenos naturales. Los futuros docentes suelen recibir una preparación matemática desarticulada de los contenidos científicos que habrán de enseñar, de modo que el conocimiento algebraico que poseen no se traduce necesariamente en decisiones de enseñanza que integren el pensamiento algebraico en el aula. Las evidencias sobre ese conocimiento no son tranquilizadoras: el análisis de docentes de educación infantil en formación reveló un nivel general insuficiente, con el conocimiento del contenido del horizonte como el más débil (Pincheira y Alsina, 2024), patrón que se reitera en

futuros docentes de primaria, donde el conocimiento común del contenido obtiene las puntuaciones más altas y el del horizonte matemático las más bajas (Pincheira y Alsina, 2026). Las entrevistas previas a una intervención con futuros maestros de primaria indicaron, asimismo, que pueden no estar preparados, en términos del conocimiento requerido, para fomentar el pensamiento algebraico (Öztürk y Işler, 2024).

El interés por este campo es creciente. Un análisis bibliométrico en Scopus identifica una atención sostenida de la comunidad investigadora al pensamiento algebraico en educación infantil y primaria, con términos como early algebra, functional thinking y generalization entre los más prominentes (Narváez et al., 2025). Desde el lado del aprendizaje, un estudio transversal con 526 estudiantes de tercero a quinto grado en Shanghái describió una progresión desde el pensamiento aritmético hasta el pensamiento algebraico simbólico (Sun et al., 2023), y la justificación en educación primaria favoreció que los niños adoptaran un lenguaje matemático cada vez más preciso y abstracto (Pinto et al., 2023). Para caracterizar los niveles de razonamiento que el docente debe reconocer y promover se ha propuesto un modelo ampliado de niveles de algebrización (Burgos et al., 2024). Estas progresiones delimitan lo que un profesor ha de saber discernir, pero no indican cómo ese saber llega a la práctica.

El tránsito del saber a la enseñanza presenta fricciones documentadas. Un porcentaje importante de profesores en formación mostró dificultades para identificar los objetos matemáticos solicitados en tareas de razonamiento algebraico (Pallauta et al., 2025), y futuros docentes que plantearon y resolvieron correctamente problemas con tablas estadísticas exhibieron limitaciones al asignar niveles de razonamiento algebraico a problemas modificados o creados por ellos mismos (Pallauta et al., 2026). Una revisión sobre la mirada profesional hacia el pensamiento algebraico de los estudiantes añade que los docentes en formación atienden bien a las características superficiales, pero interpretan y responden con menor fiabilidad, y que una visión procedimental del álgebra limita lo que notan (Melissa et al., 2026). Cuando las tareas de aprendizaje profesional se combinan con la acción de los formadores en discusiones colectivas, en

cambio, los docentes logran diferenciar y comprender el razonamiento de los estudiantes (Ribeiro et al., 2024).

A este déficit de conocimiento se suma una dimensión de creencias que la literatura latinoamericana ha comenzado a sistematizar. Una revisión sobre la autoeficacia de profesores de matemáticas en formación examinó las fuentes de eficacia y los dominios de actividad docente en los estudios disponibles (Soler y Valencia, 2025), y se han descrito perfiles que articulan la autoeficacia docente con el conocimiento especializado en la formación inicial de maestros de primaria (Avilés y Marbán, 2023). En la enseñanza del álgebra, futuros profesores de matemáticas informaron niveles de confianza de moderados a altos, con diferencias individuales notables (Özaltun y Bukova, 2026). En el contexto peruano, la evidencia sobre las percepciones de 1045 docentes de matemática acerca de su formación inicial y continua procede de la Encuesta Nacional de Docentes de 2018 (Ricaldi, 2026), lo que subraya la necesidad de modelos que vinculen conocimiento y creencias en la formación del profesorado.

La brecha que motiva este estudio es la ausencia de modelos explicativos que conecten, en un mismo sistema de ecuaciones estructurales, el conocimiento del razonamiento algebraico temprano con la integración del pensamiento algebraico en la enseñanza de las ciencias, y que pongan a prueba el papel mediador de la autoeficacia docente durante la formación inicial. La pregunta central es: ¿en qué medida el conocimiento del razonamiento algebraico temprano se asocia con la integración del pensamiento algebraico en la enseñanza de las ciencias, y qué parte de esa asociación se transmite a través de la autoeficacia docente? Se formulan cuatro hipótesis:

- **H1**, el conocimiento se relaciona positivamente con la autoeficacia docente;
- **H2**, la autoeficacia docente se relaciona positivamente con la integración;
- **H3**, el conocimiento se relaciona positivamente y de forma directa con la integración; y

- **H4**, el conocimiento ejerce un efecto indirecto positivo sobre la integración a través de la autoeficacia docente.

La contribución del estudio es teórica y metodológica, y trasciende el caso nacional. En lo teórico, separa empíricamente dos vías por las que el conocimiento disciplinar del futuro docente llega a la enseñanza —una directa y otra mediada por las creencias de capacidad— y cuantifica el peso relativo de cada una. En lo metodológico, ofrece un modelo de medición de tres constructos con validez convergente y discriminante verificada y una descomposición de efectos con intervalos bootstrap, transferible a otros sistemas de formación docente. El sistema peruano de formación inicial, en 2025, opera como contexto de contraste.

5.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

El marco principal es la teoría social cognitiva de la autoeficacia, tal como se ha aplicado a la formación inicial de profesores de matemáticas (Avilés y Marbán, 2023; Soler y Valencia, 2025). Según este enfoque, las creencias de capacidad regulan la elección de tareas, el esfuerzo invertido y la persistencia ante la dificultad, y se nutren de forma preferente de las experiencias de dominio. Aplicada a la docencia, la autoeficacia docente designa la convicción del profesor de que puede producir los resultados de aprendizaje deseados en sus estudiantes (Avilés y Marbán, 2024). Desde esta perspectiva, el conocimiento del razonamiento algebraico temprano constituye una condición necesaria pero no suficiente para su integración en la enseñanza: su movilización depende, en parte, de que el futuro docente se perciba capaz de enseñarlo.

El marco complementario es el del conocimiento didáctico del contenido, en su desarrollo como conocimiento matemático para la enseñanza del álgebra temprana (Pincheira y Alsina, 2024). Este marco distingue entre el saber disciplinar y su transformación en representaciones y tareas enseñables, y explica por qué el conocimiento algebraico puede operar también de forma directa sobre la enseñanza: quien comprende cómo se generaliza una relación funcional dispone de recursos para diseñar tareas

de modelación en ciencias, con independencia de su nivel de confianza. Ambos marcos describen, por tanto, procesos que no se excluyen entre sí, uno centrado en las creencias que regulan la acción y otro en el saber que la hace posible.

La derivación hacia las hipótesis es la siguiente.

- Si la autoeficacia se nutre de la experiencia de dominio disciplinar, un mayor conocimiento del razonamiento algebraico temprano debería asociarse con mayor autoeficacia docente (H1).
- Si la autoeficacia regula la acción pedagógica, una mayor autoeficacia debería asociarse con mayor integración del pensamiento algebraico en la enseñanza de las ciencias (H2).
- Si el conocimiento disciplinar se convierte también en decisiones didácticas por una vía propia, debe observarse un efecto directo del conocimiento sobre la integración aun controlando la autoeficacia (H3).
- Y si la autoeficacia transmite parte del efecto del conocimiento, debe observarse un efecto indirecto distinto de cero (H4).

La coexistencia de H3 y H4 configura una mediación parcial, forma que cabe esperar cuando ambos marcos describen procesos complementarios y no excluyentes, y no una mediación completa, que exigiría que el conocimiento actuara únicamente a través de las creencias de capacidad.

5.3 MATERIALES Y MÉTODOS

5.3.1 Diseño y participantes

El estudio adoptó un diseño no experimental, transversal y explicativo, con recolección de datos en un solo momento durante 2025. La muestra estuvo conformada por N = 396 estudiantes de formación inicial docente en la especialidad de ciencias de instituciones de educación superior pedagógica y universitaria del Perú, seleccionados mediante muestreo no

probabilístico por conveniencia; por ello no se afirma la representatividad de la población nacional. Se incluyó a quienes habían completado o cursaban asignaturas de contenido matemático y de didáctica de las ciencias, y se excluyeron los protocolos incompletos o con patrones de respuesta inválidos.

El tamaño muestral se justificó con una potencia a priori para cada análisis, y N = 396 cubre el más exigente. Para el modelo estructural se aplicó la prueba de ajuste cercano basada en el RMSEA (MacCallum et al., 1996), que contrasta H0: $\varepsilon \leq .05$ frente a H1: $\varepsilon = .08$ con $\alpha = .05$; con 9 indicadores y 21 parámetros libres, gl = 24, y N = 375 produce $\lambda 0 = 22.44$, $\lambda 1 = 57.45$, χ^2 crítico = 67.14 y una potencia de .801, de modo que se requieren al menos 375 participantes. Para la regresión de la variable de resultado sobre dos predictores se empleó la prueba F de un coeficiente individual (Cohen, 1988), con el tamaño de efecto anclado en estudios publicados: Aderinoye et al. (2025) informan r = .18 y .35 con $\beta = .26$ y .40, de donde se recupera $f^2 = .080$ y .189, y Razali et al. (2022) y Hayati et al. (2023) informan correlaciones de .356 y .543, equivalentes a $f^2 = .145$ y .418. Se planificó sobre la cota inferior, $f^2 = .080$, que exige N = 102 (potencia = .804); con N = 396 la potencia supera .999. El margen sobre el mínimo de 375 es de 21 participantes.

5.3.2 Instrumentos y variables

Las tres variables latentes se midieron con escalas de tres ítems tipo Likert de cinco puntos, operacionalizadas en la Tabla 5-1. Las escalas de conocimiento del razonamiento algebraico temprano y de integración son de construcción propia, con un ítem por cada componente del dominio, y su estructura se verifica en el modelo de medición. La de autoeficacia docente adapta la escala de Avilés y Marbán (2024): conserva su formulación en términos de capacidad percibida para lograr aprendizajes y reorienta los ítems hacia la enseñanza del razonamiento algebraico en las clases de ciencias. Se registraron, además, dos covariables de control: las horas semanales de estudio autónomo de matemática, por autoinforme, y el sexo, dicotómico.

Tabla 5-1

Operacionalización de los constructos del modelo

Variable	Rol en el modelo	Tipo de medición	Ítems	Escala	Contenido medido
Conocimiento del razonamiento algebraico temprano	Exógena (X)	Latente reflectiva	3	Likert de 5 puntos	Generalización de patrones, reconocimiento del pensamiento algebraico, relaciones funcionales simbolizadas (escala de construcción propia; estructura verificada en la Tabla 5-3)
Autoeficacia docente	Mediadora (M)				Confianza para enseñar el razonamiento algebraico en clases de ciencias (adaptación de Avilés y Marbán, 2024)
Integración del pensamiento algebraico en la enseñanza de las ciencias	Endógena final (Y)				Modelación simbólica de fenómenos, tareas de generalización con datos, vínculo de patrones con conceptos científicos (escala de construcción propia; estructura verificada en la Tabla 5-3)
Horas semanales de estudio autónomo de matemática	Covariable de controll	Observada continua	1	Horas por semana	Autoinforme
Sexo		Observada dicotómica		0/1	Registro sociodemográfico

Nota: Las covariables de control se emplean en los modelos de regresión de robustez (Tabla 5-8); el modelo de ecuaciones estructurales contiene únicamente las tres variables latentes..

5.3.3 Análisis de datos

El análisis siguió dos etapas.

- En la primera se evaluó el modelo de medición mediante análisis factorial confirmatorio: cargas estandarizadas (λ), comunalidades (R^2), alfa de Cronbach (α), omega de McDonald (ω), varianza extraída media (AVE) y validez discriminante por el criterio de Fornell y Larcker

(1981) y la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015).

- En la segunda se estimó por máxima verosimilitud (semopy) un modelo de ecuaciones estructurales basado en covarianzas, enfoque también empleado en estudios peruanos sobre autoeficacia docente (Velez et al., 2024), con las rutas X → M, M → Y y X → Y.

El efecto indirecto a × b y el efecto total se evaluaron con intervalos de confianza al 95 % por bootstrap de 5000 réplicas, significativos cuando excluían el cero. El ajuste global se describió con χ^2, gl, χ^2/gl, CFI, TLI, RMSEA y SRMR, bajo los puntos de corte de Hu y Bentler (1999), y se comparó con un modelo alternativo de un solo factor mediante la diferencia de χ^2. La robustez se examinó con una regresión jerárquica por mínimos cuadrados ordinarios con errores estándar HC3 sobre puntuaciones compuestas, que incorpora las covariables, y con pruebas de estrés del coeficiente focal (jackknife, submuestreo, remuestreo bootstrap, winsorización y predictor placebo). Como limitación del diseño, el modelo estimado es de un solo nivel y no modela la dependencia entre participantes de una misma institución.

5.4 RESULTADOS

5.4.1 Estadísticos descriptivos y correlaciones

La Tabla 5-2 presenta los estadísticos descriptivos y las correlaciones. Las tres variables latentes se correlacionaron de forma positiva y moderada (r entre .541 y .572). Las horas semanales de estudio autónomo de matemática se asociaron levemente con la integración ($r = .108$, $p < .05$), el sexo no se asoció significativamente con ninguna variable, y las distribuciones de las variables latentes fueron aproximadamente simétricas (asimetría y curtosis en valor absoluto menores que .15).

Tabla 5-2

Estadísticos descriptivos y correlaciones entre constructos

	Variable	M	DE	Asimetría	Curtosis	1	2	3	4
1	Conocimiento del razonamiento algebraico temprano	3.00	0.66	−0.05	−0.04	—			
2	Integración del pensamiento algebraico en la enseñanza de las ciencias	3.00	0.67	−0.04	−0.15	.572*** [.502, .635]	—		
3	Autoeficacia docente	3.00	0.67	−0.08	−0.04	.569*** [.498, .632]	.541*** [.467, .607]	—	
4	Horas semanales de estudio autónomo de matemática	5.91	4.47	1.47	1.82	.064 [−.035, .161]	.108* [.009, .204]	.024 [−.075, .122]	—
5	Sexo	0.49	0.50	0.03	−2.00	−.089 [−.186, .010]	−.067 [−.165, .032]	.017 [−.082, .115]	.017 [−.082, .115]

Nota: N = 396. Entre corchetes, intervalos de confianza al 95 % por transformación Z de Fisher. Las medias y desviaciones estándar de las variables latentes se expresan en la escala de respuesta de cinco puntos. Sexo, variable dicotómica (0/1). Las correlaciones del sexo con la autoeficacia docente (r = .0170) y con las horas semanales de estudio autónomo de matemática (r = .0166) son distintas y coinciden solo al redondear a tres decimales, por lo que comparten también su intervalo de confianza redondeado. *p < .05; **p < .01; ***p < .001.

5.4.2 Modelo de medición

Las nueve cargas estandarizadas oscilaron entre .711 y .876, todas por encima de .70 (Tabla 5-3). La fiabilidad fue adecuada en los tres constructos (α entre .800 y .866; ω entre .802 y .865) y la AVE superó .50 en todos ellos (entre .574 y .682).

Tabla 5-3

Modelo de medición: cargas, fiabilidad y validez convergente

Constructo	Ítem	λ	R^2	α	ω	AVE
Conocimiento del razonamiento algebraico temprano	Identifico patrones numéricos y los generalizo mediante expresiones algebraicas	.802	.643	.800	.802	.574
	Reconozco el pensamiento algebraico en tareas de generalización de estudiantes	.758	.575			
	Traduzco relaciones funcionales entre magnitudes a representaciones simbólicas	.711	.506			
Autoeficacia docente	Identifico patrones numéricos y los generalizo mediante expresiones algebraicas	.802	.643	.800	.802	.574
	Reconozco el pensamiento algebraico en tareas de generalización de estudiantes	.758	.575			
	Traduzco relaciones funcionales entre magnitudes a representaciones simbólicas	.711	.506			
Integración del pensamiento algebraico en la enseñanza de las ciencias	Identifico patrones numéricos y los generalizo mediante expresiones algebraicas	.802	.643	.800	.802	.574
	Reconozco el pensamiento algebraico en tareas de generalización de estudiantes	.758	.575			
	Traduzco relaciones funcionales entre magnitudes a representaciones simbólicas	.711	.506			

Nota: N = 396. λ = carga factorial estandarizada del análisis factorial confirmatorio; R^2 = comunalidad del ítem; α = alfa de Cronbach; ω = omega de McDonald; AVE = varianza extraída media. Valores de referencia: λ ≥ .70, α y ω ≥ .70, AVE ≥ .50.

Según la Tabla 5-4, la raíz cuadrada de la AVE de cada constructo superó su mayor correlación con los demás, y todos los valores HTMT fueron menores que .85, con un máximo de .681 para el par conocimiento–autoeficacia.

Tabla 5-4

Validez discriminante: criterio de Fornell-Larcker y razón HTMT

Constructo		1	2	3
1.	Conocimiento del razonamiento algebraico temprano	**.758**	.681	.571
2.	Autoeficacia docente	.569	**.826**	.551
3.	Integración del pensamiento algebraico en la enseñanza de las ciencias	.572	.541	**.819**

Nota: N = 396. En la diagonal (negrita), raíz cuadrada de la AVE; bajo la diagonal, correlaciones entre constructos (Tabla 2); sobre la diagonal, razón heterorrasgo-monorrasgo (HTMT). Criterios: √AVE mayor que las correlaciones de su fila y columna; HTMT < .85.

5.4.3 Ajuste global del modelo

El modelo hipotetizado de tres factores presentó un χ^2 no significativo, $\chi^2(24) = 29.41$, $p = .205$, con índices que cumplieron los puntos de corte de referencia (Tabla 5-5). El modelo alternativo de un solo factor ajustó de forma claramente inferior, y la diferencia entre ambos fue significativa, $\Delta\chi^2(3) = 490.61$, $p < .001$. La Figura 5-1 reúne las evidencias de validez del modelo de medición y de ajuste.

Tabla 5-5

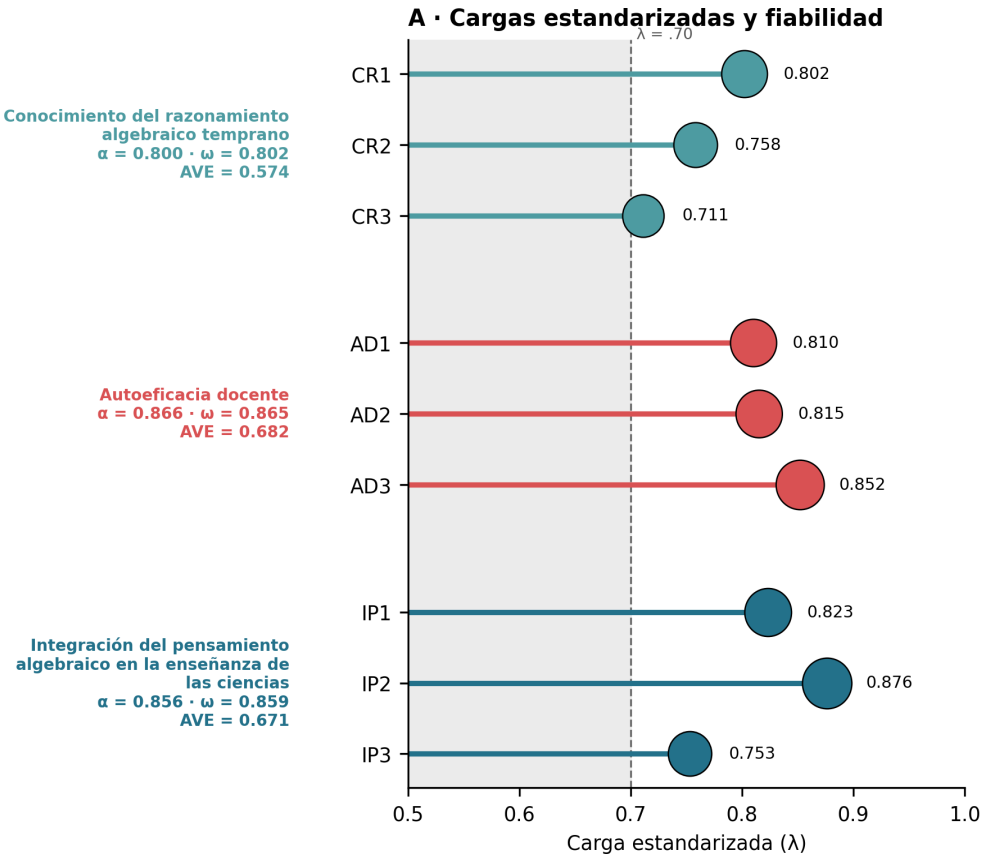
Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor

Índice	Modelo	
	hipotetizado (tres factores)	alternativo (un factor)
χ^2	29.41	520.02
gl	24	27
p (χ^2)	.205	< .001
χ^2/gl	1.23	19.26
CFI	.997	.723
TLI	.995	.631
RMSEA	.024	.215
SRMR	.022	.134

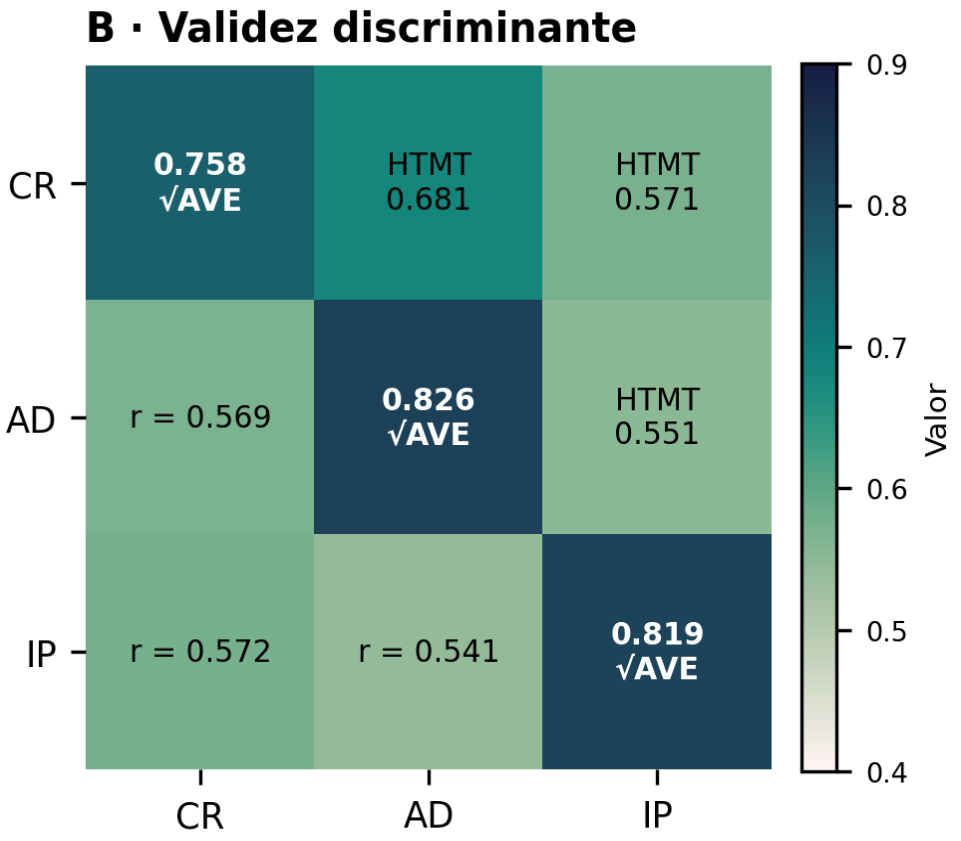
Nota: N = 396. Estimación por máxima verosimilitud. Diferencia entre modelos: $\Delta\chi^2(3) = 490.61$, $p < .001$. Puntos de corte de referencia: CFI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu y Bentler, 1999).

Figura 5-1

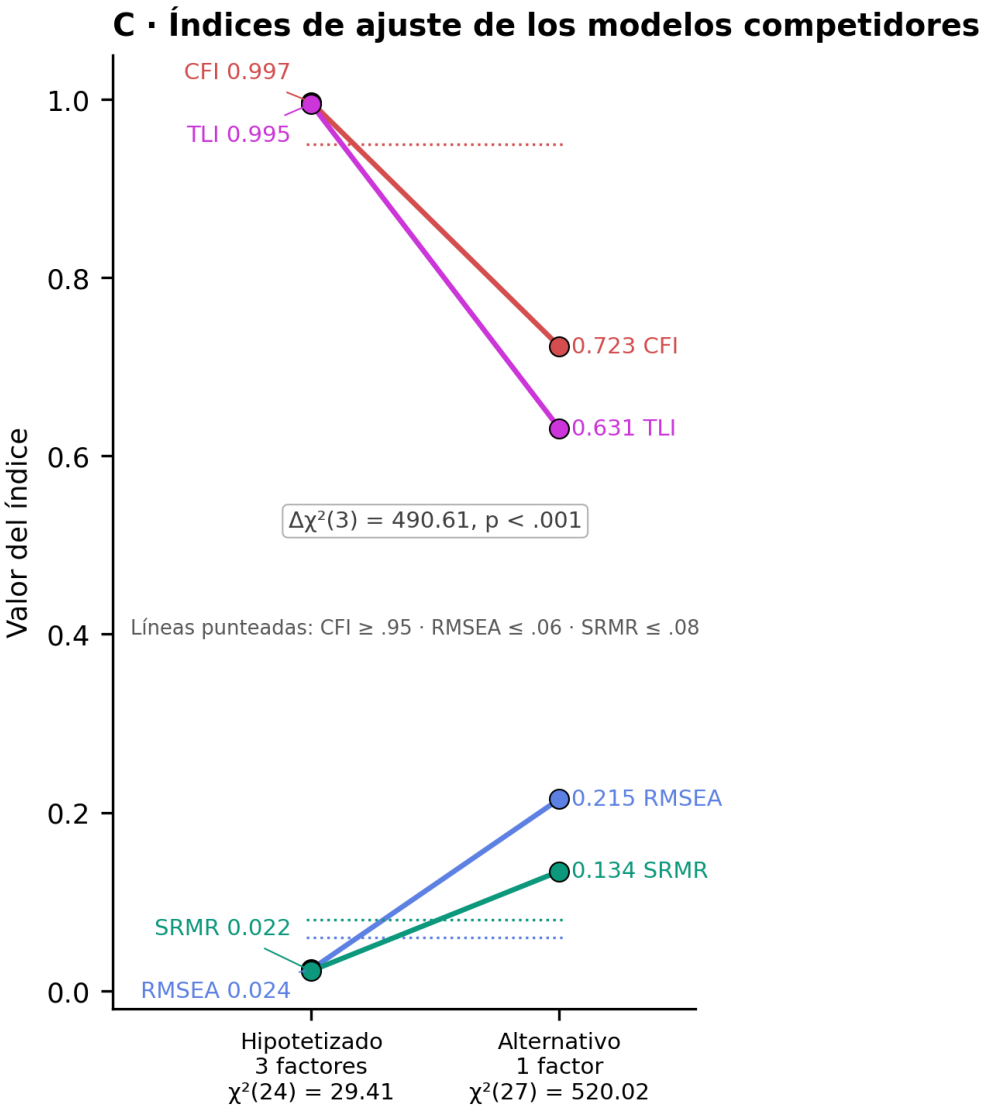
Evidencia de validez del modelo de medición y de ajuste global



Nota. N = 396. Panel A: cargas estandarizadas de los nueve ítems (el tamaño del marcador es proporcional a la comunalidad R^2; zona sombreada, cargas menores que .70), con alfa de Cronbach, omega de McDonald y varianza extraída media de cada constructo: conocimiento del razonamiento algebraico temprano (α = .800, ω = .802, AVE = .574), autoeficacia docente (α = .866, ω = .865, AVE = .682) e integración del pensamiento algebraico en la enseñanza de las ciencias (α = .856, ω = .859, AVE = .671). Panel B: matriz de validez discriminante con la raíz cuadrada de la AVE en la diagonal (.758, .826 y .819), las correlaciones entre constructos bajo la diagonal (.569, .572 y .541) y la razón heterorrasgo-monorrasgo sobre ella (.681, .571 y .551; máximo .681 < .85).



Nota. N = 396. Panel B: matriz de validez discriminante con la raíz cuadrada de la AVE en la diagonal (.758, .826 y .819), las correlaciones entre constructos bajo la diagonal (.569, .572 y .541) y la razón heterorrasgo-monorrasgo sobre ella (.681, .571 y .551; máximo .681 < .85).



Nota. N = 396. Panel C: índices de ajuste del modelo hipotetizado de tres factores ($\chi^2(24) = 29.41$) frente al modelo alternativo de un factor ($\chi^2(27) = 520.02$): CFI .997 frente a .723, TLI .995 frente a .631, RMSEA .024 frente a .215 y SRMR .022 frente a .134; $\Delta\chi^2(3) = 490.61$, $p < .001$. Las líneas punteadas marcan los puntos de corte de referencia CFI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu y Bentler, 1999).

5.4.4 Modelo estructural y descomposición de efectos

Las tres rutas estructurales directas resultaron positivas y significativas (Tabla 5-6), con lo que H1, H2 y H3 se respaldaron. El modelo explicó el 43.6 % de la varianza de la autoeficacia y el 31.2 % de la varianza de la integración.

Tabla 5-6

Rutas estructurales directas del modelo

Hipótesis	Ruta	β	p	IC 95 % bootstrap	Decisión
H1	Conocimiento del razonamiento algebraico temprano → Autoeficacia docente	.660	< .001	[.578, .737]	Se respalda
H2	Autoeficacia docente → Integración del pensamiento algebraico en la enseñanza de las ciencias	.317	< .001	[.148, .489]	Se respalda
H3	Conocimiento del razonamiento algebraico temprano → Integración del pensamiento algebraico en la enseñanza de las ciencias	.296	< .001	[.118, .467]	Se respalda

Nota: N = 396. β = coeficiente estandarizado del modelo estructural latente (máxima verosimilitud); la incertidumbre se expresa con intervalos de confianza por bootstrap de 5000 réplicas sobre el coeficiente estandarizado. Varianza explicada, derivada de la solución estandarizada: R^2 = .436 para la autoeficacia docente y R^2 = .312 para la integración del pensamiento algebraico en la enseñanza de las ciencias.

El efecto indirecto fue positivo y su intervalo excluyó el cero, con lo que H4 se respaldó (Tabla 5-7). Al ser también significativo el efecto directo, la mediación fue parcial: el 41.4 % del efecto total se transmitió por la autoeficacia docente. La Figura 5-2 integra el modelo y la descomposición.

Tabla 5-7

Descomposición del efecto del conocimiento del razonamiento algebraico temprano sobre la integración

Efecto	Composición	Estimación	IC 95 % bootstrap	Proporción del efecto total
Indirecto (H4)	a × b = .660 × .317	.209	[.095, .335]	41.4 %
Directo (c′)	X → Y controlando M	.296	[.118, .467]	58.6 %
Total (c)	c′ + a × b	.505	[.400, .605]	100 %

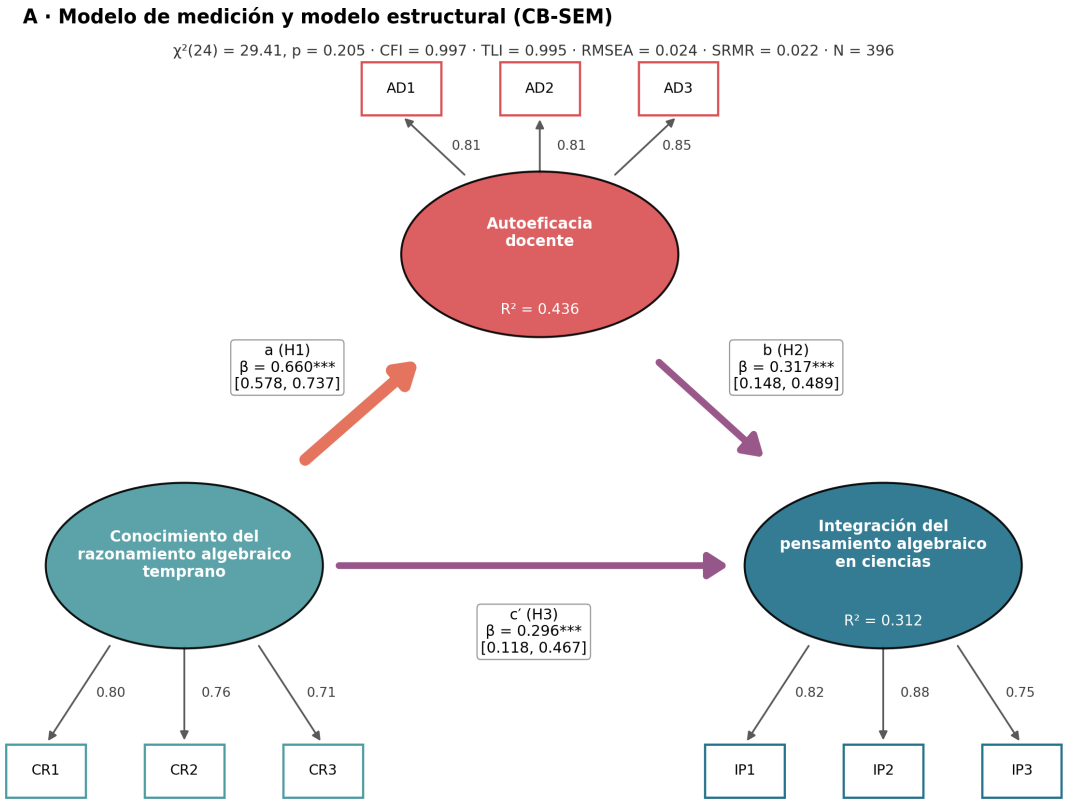
Nota: N = 396. Coeficientes estandarizados; intervalos por bootstrap de 5000 réplicas (5000 convergidas). Un efecto se considera significativo cuando su intervalo excluye el cero. La coexistencia de efecto directo e indirecto significativos indica mediación parcial.

5.4.5 Robustez

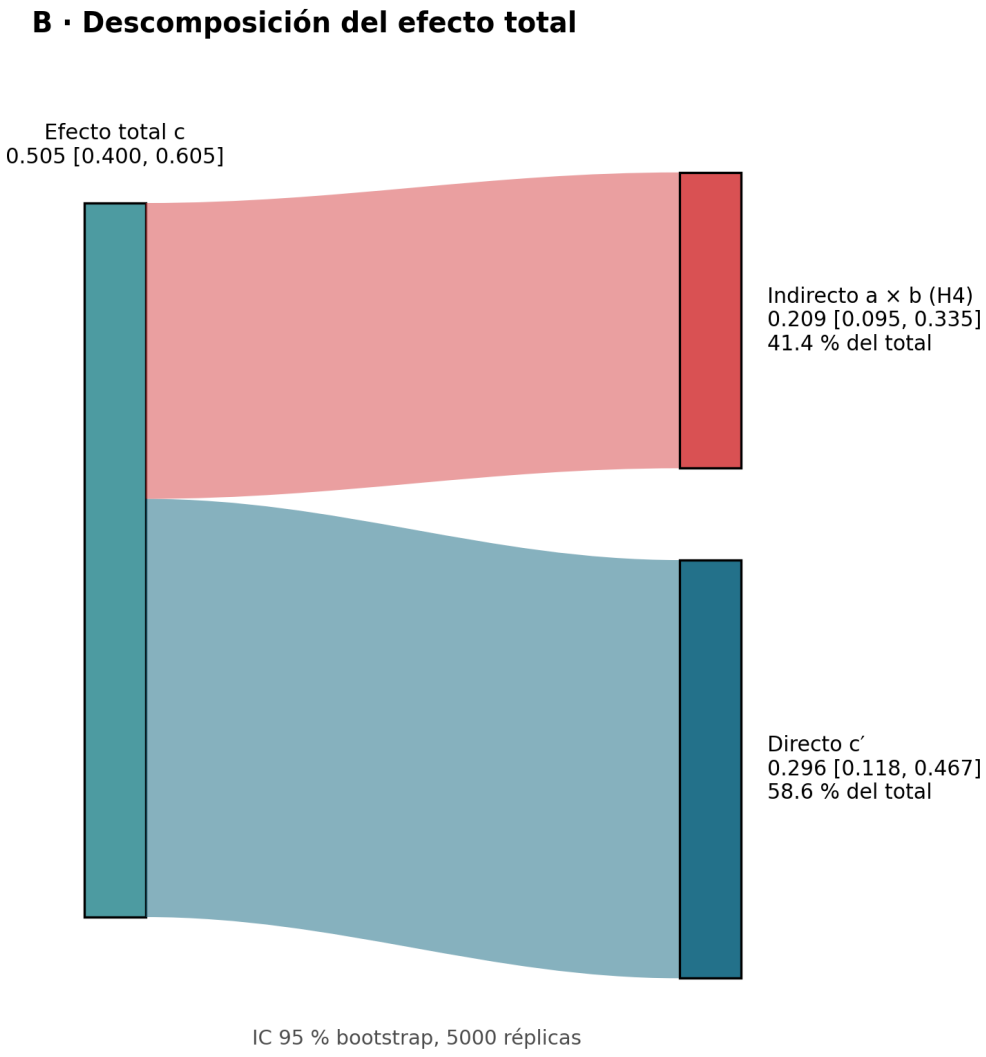
La regresión jerárquica sobre las puntuaciones compuestas, con las covariables incorporadas, reprodujo la dirección y la significancia de las relaciones del modelo estructural (Tabla 5-8). El conocimiento añadió un $\Delta R^2 = .317$ sobre las covariables, la autoeficacia un $\Delta R^2 = .070$ adicional, el coeficiente del conocimiento disminuyó de .479 a .322 al incorporarla, las covariables no alcanzaron significación en el modelo completo y el factor de inflación de la varianza máximo fue 1.01.

Figura 5-2

Modelo covariante estructural y de medición para la descomposición del efecto del conocimiento del razonamiento algebraico temprano en la integración del pensamiento algebraico dentro de la enseñanza de las ciencias



Nota. N = 396; estimación por máxima verosimilitud. Panel A: variables latentes (elipses) con sus tres indicadores (rectángulos; CR = conocimiento del razonamiento algebraico temprano, AD = autoeficacia docente, IP = integración del pensamiento algebraico en ciencias) y cargas estandarizadas redondeadas a dos decimales (valores exactos en la Tabla 3); el grosor de cada ruta es proporcional a su coeficiente estandarizado, con IC 95 % por bootstrap de 5000 réplicas: a = .660 [.578, .737], b = .317 [.148, .489] y c′ = .296 [.118, .467], todos con $p < .001$; $R^2 = .436$ para la autoeficacia docente y $R^2 = .312$ para la integración; ajuste: $\chi^2(24) = 29.41$, $p = .205$, CFI = .997, TLI = .995, RMSEA = .024, SRMR = .022.



Nota. N = 396; estimación por máxima verosimilitud. Panel B: flujo del efecto total c = .505 [.400, .605] repartido en efecto indirecto a × b = .209 [.095, .335] (41.4 %) y efecto directo c′ = .296 [.118, .467] (58.6 %); la altura de cada banda es proporcional a su estimación. *** p < .001.

Tabla 5-8

Regresión jerárquica de la integración del pensamiento algebraico en la enseñanza de las ciencias

Predictor	M1: covariables	M2: + conocimiento	M3: + autoeficacia
Horas semanales de estudio autónomo de matemática	0.021* (0.010) [0.002, 0.040]	0.014 (0.008) [–0.002, 0.029]	0.015 (0.008) [–0.001, 0.029]
Sexo	–0.117 (0.085) [–0.284, 0.050]	–0.031 (0.070) [–0.168, 0.107]	–0.068 (0.066) [–0.198, 0.062]
Conocimiento del razonamiento algebraico temprano	—	0.479*** (0.037) [0.407, 0.551]	0.322*** (0.044) [0.236, 0.407]
Autoeficacia docente	—	—	0.388*** (0.060) [0.270, 0.505]
R^2	.016	.333	.403
R^2 ajustado	.011	.328	.397
F global	F(2, 393) = 3.27, p = .039	F(3, 392) = 59.00, p < .001	F(4, 391) = 59.13, p < .001
ΔR^2	—	.317	.070
F de cambio	—	F(1, 392) = 186.07, p < .001	F(1, 391) = 45.99, p < .001
f^2 del bloque añadido	—	.475	.118

Nota: N = 396. Coeficientes de mínimos cuadrados ordinarios sobre puntuaciones compuestas, con errores estándar robustos HC3 entre paréntesis e intervalos de confianza al 95 % entre corchetes. $f^2 = (R^2_{incluido} - R^2_{excluido})/(1 - R^2_{incluido})$; el f^2 del modelo completo frente a M1 es .648. *p < .05; **p < .01; ***p < .001.

Las pruebas de estrés del coeficiente del conocimiento en M2 (Tabla 5-9) mostraron que conservó el signo y la significancia en el 100 % de las remuestras del jackknife, del submuestreo al 80 % y al 60 % y del remuestreo bootstrap, se mantuvo tras winsorizar la variable de resultado y no resultó significativo cuando el predictor se sustituyó por un placebo aleatorio. La Figura 5-3 resume la relación bivariada, la varianza explicada por bloques y esta estabilidad.

Tabla 5-9

Pruebas de estrés del coeficiente del conocimiento del razonamiento algebraico temprano (modelo M2)

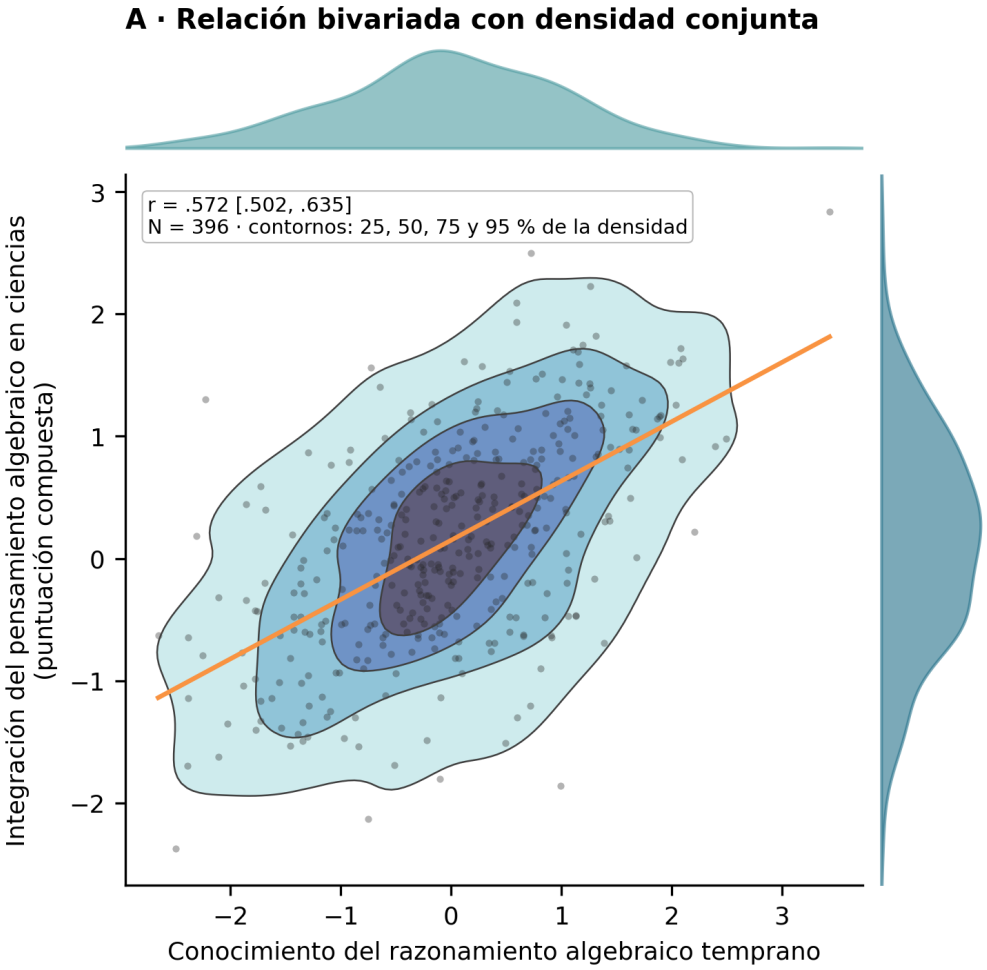
Prueba	Réplicas	β medio	DE	IC 95 %	Signo estable (%)	Significativo (%)
Jackknife (dejando uno fuera)	396	0.479	0.002	[0.475, 0.483]	100 %	100 %
Submuestreo al 80 %	100	0.481	0.018	[0.445, 0.515]	100 %	100 %
Submuestreo al 60 %	100	0.483	0.029	[0.436, 0.534]	100 %	100 %
Remuestreo bootstrap	1000	0.478	0.036	[0.406, 0.551]	100 %	100 %
Variable de resultado winsorizada al 5 %	1	0.440	—	[0.376, 0.503]	Sí	Sí ($p < .001$)
Predictor placebo aleatorio	1	–0.057	—	[–0.145, 0.031]	—	No ($p = .203$)

Nota: N = 396. Coeficiente de mínimos cuadrados ordinarios del conocimiento del razonamiento algebraico temprano en el modelo M2 (β base = 0.479, $p < .001$). «Signo estable» y «Significativo» indican el porcentaje de re-muestras que conservan el signo positivo y $p < .05$. La prueba del placebo es un control negativo: un predictor sin relación con la variable de resultado no debe producir un efecto significativo.

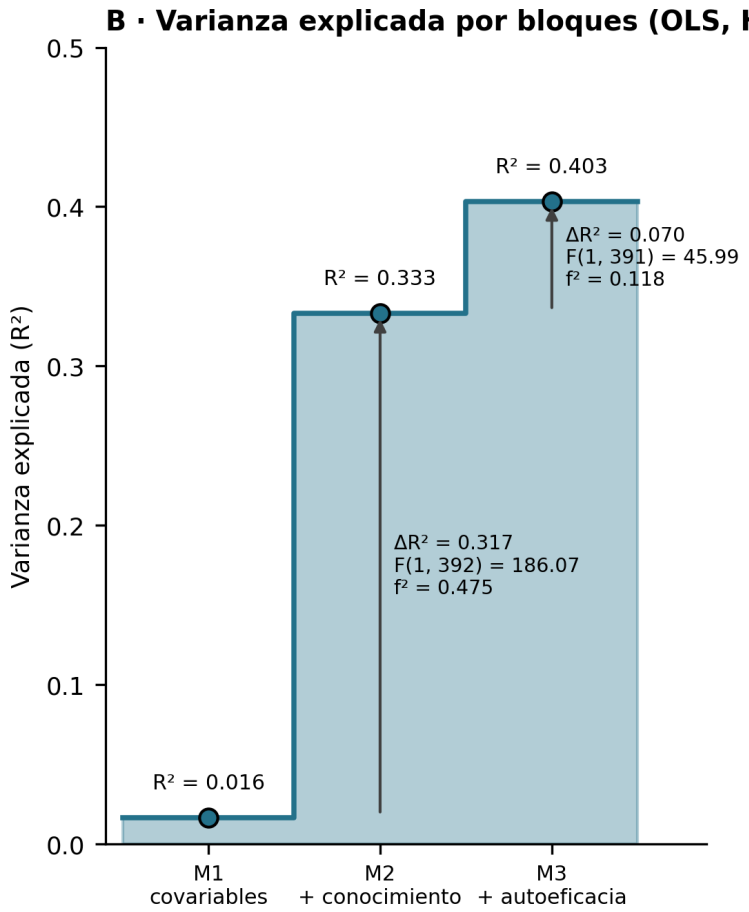
La Tabla 5-10 resume la correspondencia entre cada hipótesis, el contraste aplicado y el estadístico que la decide.

Figura 5-3

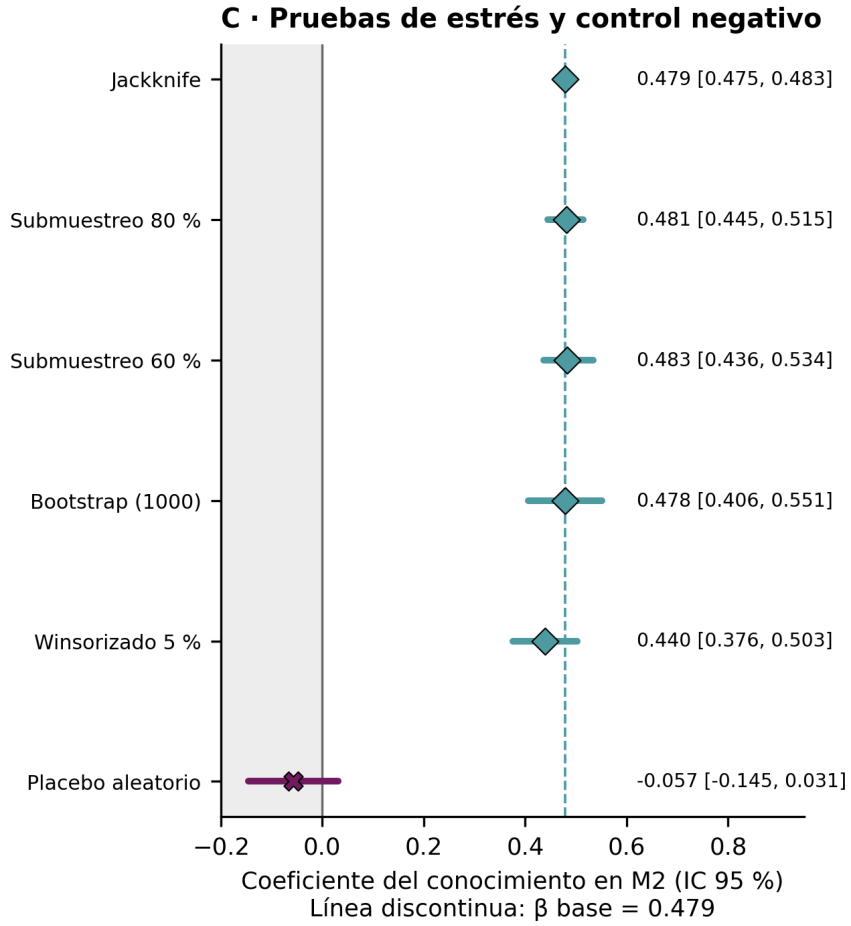
Relación bivariada, varianza explicada por bloques y estabilidad del coeficiente del conocimiento del razonamiento algebraico temprano



Nota. N = 396; puntuaciones compuestas. Panel A: dispersión del conocimiento del razonamiento algebraico temprano (puntuación estandarizada) y de la integración del pensamiento algebraico en la enseñanza de las ciencias, con contornos de densidad conjunta que encierran el 25, 50, 75 y 95 % de los casos, recta de mínimos cuadrados y densidades marginales; r = .572 [.502, .635].



Nota. N = 396; puntuaciones compuestas. Panel B: R^2 de la regresión jerárquica con errores estándar HC3: M1 con covariables, $R^2 = .016$; M2 con el conocimiento, $R^2 = .333$ ($\Delta R^2 = .317$, $F(1, 392) = 186.07$, $p < .001$, $f^2 = .475$); M3 con la autoeficacia docente, $R^2 = .403$ ($\Delta R^2 = .070$, $F(1, 391) = 45.99$, $p < .001$, $f^2 = .118$).



Nota. N = 396; puntuaciones compuestas. Panel C: coeficiente del conocimiento en M2 con IC 95 % bajo jackknife (0.479 [0.475, 0.483]), submuestreo al 80 % (0.481 [0.445, 0.515]) y al 60 % (0.483 [0.436, 0.534]), remuestreo bootstrap de 1000 réplicas (0.478 [0.406, 0.551]) y winsorización al 5 % (0.440 [0.376, 0.503]), frente al control negativo con un predictor placebo aleatorio (−0.057 [−0.145, 0.031], p = .203); línea discontinua, coeficiente base 0.479.

Tabla 5-10

Trazabilidad de las hipótesis: contraste, procedimiento y decisión

Hipótesis	Contraste	Procedimiento	Estadístico de decisión	Resultado	Tabla
H1	Conocimiento → autoeficacia (+)	Ruta a del modelo estructural (máxima verosimilitud)	β con IC 95 % bootstrap	β = .660 [.578, .737]; se respalda	5-6
H2	Autoeficacia → integración (+)	Ruta b del modelo estructural	β con IC 95 % bootstrap	β = .317 [.148, .489]; se respalda	5-6
H3	Conocimiento → integración, efecto directo (+)	Ruta c′ controlando la autoeficacia	β con IC 95 % bootstrap	β = .296 [.118, .467]; se respalda	5-6
H4	Efecto indirecto vía autoeficacia (+)	Producto a × b con bootstrap de 5000 réplicas	IC 95 % que excluye el cero	a × b = .209 [.095, .335]; se respalda, mediación parcial	5-7

Nota: N = 396. Las regresiones jerárquicas (Tabla 5-8) y las pruebas de estrés (Tabla 5-9) son análisis de robustez sobre puntuaciones compuestas y no deciden ninguna hipótesis..

5.5 DISCUSIÓN

El hallazgo más desconcertante no es la mediación esperada, sino que la mayor parte del efecto total del conocimiento sobre la integración —el 58.6 % (Tabla 5-7)— no transitó por la autoeficacia, pese a que la ruta entre conocimiento y autoeficacia fue la más intensa del modelo (Tabla 5-6). Si las creencias de capacidad fueran el conducto principal por el que el saber disciplinar llega al aula, cabría esperar una mediación casi completa; lo observado sugiere, en cambio, que el futuro docente que domina la generalización de patrones y la simbolización de relaciones funcionales dispone de recursos para planificar tareas de modelación que no dependen de cuán capaz se perciba. La lectura coherente con los cuatro contrastes (Tabla 5-10) es una mediación parcial en la que dos procesos operan en paralelo y ninguno absorbe al otro. La medición sostiene esa

lectura: los tres constructos son empíricamente distinguibles y el modelo de tres factores ajusta de forma muy superior al de uno solo (Tablas 3 a 5), de modo que conocimiento, autoeficacia e integración no parecen expresiones de una misma disposición general.

Este patrón contrasta con la evidencia de que, en docentes en servicio, las creencias de eficacia pesan más que el conocimiento. En una muestra de profesores de ciencias, matemática y lengua de secundaria baja en Estonia, el análisis de conglomerados mostró que las creencias de autoeficacia parecían más importantes que el conocimiento pedagógico general para la calidad instruccional, aunque este último lo era más para la persistencia profesional (Leijen et al., 2024). Nuestros datos no replican esa primacía: la vía directa del conocimiento resultó de magnitud comparable a la mediada. Varias diferencias pueden explicarlo, y ninguna puede dirimirse con este diseño. El conocimiento evaluado aquí es específico del dominio algebraico y no pedagógico general; los participantes carecen aún de experiencia docente sostenida, de modo que la autoeficacia no ha sido contrastada con la práctica; y la técnica analítica difiere, pues una tipología de perfiles puede resaltar combinaciones extremas que un modelo de rutas lineales promedia. Desde una perspectiva crítica, la divergencia sugiere que el peso relativo de conocimiento y creencias no es una constante, sino que depende de la etapa de carrera y del grado de especificidad del conocimiento medido.

Más llamativo todavía es el contraste con el trabajo de Lumbantoruan (2023), quien, con futuros docentes de primaria, halló que las oportunidades de aprendizaje y las creencias sobre las expectativas de resultados se relacionaban con el conocimiento matemático para la enseñanza, mientras que la eficacia personal para enseñar matemática no mostraba relación alguna ($\beta = 0.017$). Frente a esa nulidad, la asociación intensa que observamos entre conocimiento y autoeficacia exige cautela interpretativa. Tres aspectos separan ambos estudios: la dirección en que se especifica la relación, ya que allí la eficacia se trata como predictora del conocimiento y aquí como consecuencia de este; el carácter general o específico del dominio en que se mide la eficacia, puesto que nuestra escala

se reorienta hacia la enseñanza del razonamiento algebraico en ciencias; y la naturaleza de la medida de conocimiento, declarada en nuestro caso. Es plausible que la autoeficacia genérica para enseñar matemática se desvincule del saber, mientras que una autoeficacia anclada en tareas algebraicas concretas se alimente de él, aunque esta conjetura requiere contrastes directos.

La sensibilidad al método de medición obliga a una lectura prudente de la ruta entre autoeficacia e integración. El estudio de vídeo de TALIS mostró que el vínculo entre autoeficacia docente y calidad de la enseñanza es grande cuando la calidad la informa el propio profesor, pero inexistente cuando la califican observadores expertos, y que prácticamente no hay relación con los resultados de los estudiantes (Jerrim et al., 2024). Como en nuestro modelo tanto la autoeficacia como la integración proceden del autoinforme, parte de la covariación entre ambas puede deberse a varianza compartida del método, y la integración debe leerse como práctica declarada y no observada. Esta advertencia afecta con más fuerza a la ruta b que a la ruta directa del conocimiento, aunque ambas se estiman con el mismo instrumento; por ello, la proporción atribuida a la vía mediada podría estar sobrestimada, y la comparación entre vías es más fiable en su ordenamiento general que en sus cifras exactas.

La coincidencia con la literatura sobre la génesis de la autoeficacia es, por contraste, más directa. Los futuros docentes de primaria incrementaron su autoeficacia para la enseñanza STEM integrada al cursar asignaturas de métodos, y entre las fuentes que emergieron figuraron las experiencias de dominio, la activación emocional, las experiencias vicarias y la persuasión verbal (Menon et al., 2023; Menon et al., 2025), con una reducción simultánea de los desafíos percibidos (Menon et al., 2026). Un patrón análogo se describió en un curso semestral de ciencias basado en indagación con futuros maestros (Ibourk y Mathis, 2024). Esta línea es consistente con la ruta entre conocimiento y autoeficacia, si se acepta que dominar el contenido constituye una experiencia de dominio.

Las condiciones del entorno matizan, no obstante, la transferencia de esas creencias a la práctica. La influencia de las creencias, las actitudes, la autoeficacia y el contexto escolar sobre las prácticas de enseñanza STEM integrada en profesorado de ciencias (Shahali y Halim, 2023), y la mayor dependencia del contexto que declaran los futuros maestros formados en STEM integrada, junto con una menor intención de adoptar el enfoque (Lupión et al., 2026), sugieren que la autoeficacia es una condición necesaria pero no suficiente para la integración. Esto es consistente con el modesto peso de la ruta b y con la existencia de una vía directa, y señala además un límite de nuestro modelo de un solo nivel, que no captura el contexto institucional.

La complementariedad de los dos marcos teóricos encuentra apoyo adicional en estudios sobre conocimiento algebraico docente. Futuros maestros cuya resolución de un problema de generalización de patrones se relaciona con las dificultades que anticipan para enseñarlo (Lee y Lee, 2023) ilustran cómo el saber disciplinar modela decisiones didácticas por sí mismo, y los procesos típicamente algebraicos que movilizan docentes en servicio —particularización-generalización, representación-significación, descomposición-reificación y modelación— apuntan en la misma dirección (Gaita et al., 2024). La vía mediada, en cambio, se conecta con la dimensión afectiva: la percepción positiva de la matemática en la formación inicial del docente de primaria se asocia de forma muy fuerte con el rendimiento (Rho = 0,829), y las experiencias adversas con docentes la deterioran (López et al., 2026). Que ninguna vía absorba a la otra sugiere que conocimiento y creencias de capacidad se refuerzan más que se sustituyen.

Las implicaciones para la formación inicial se derivan de esa doble vía, con la cautela propia de un diseño transversal. Dado que el dominio del contenido aparece como fuente sustantiva de autoeficacia antes de la experiencia docente sostenida, resultan prometedoras las experiencias que combinan aprendizaje del contenido con ensayo de la enseñanza. Un club de vídeos de siete sesiones para mirar profesionalmente el pensamiento algebraico de los estudiantes (Pinto y Cortés, 2025), el uso de contextos de

divisibilidad como alternativa para desarrollar el razonamiento algebraico elemental (Gaita et al., 2026) y la creación de problemas de álgebra temprana por docentes de primaria, que permite caracterizar su pensamiento matemático (Hermosa y Sánchez, 2024), son formatos compatibles con ese doble propósito, aunque el presente estudio no los evaluó. En términos de evaluación formativa, la medición de la autoeficacia podría servir como indicador de cuánto del conocimiento adquirido se percibe transferible a la enseñanza, siempre que se complemente con medidas observadas de la práctica.

Por último, la robustez de los resultados descansa en evidencias convergentes obtenidas con estimadores distintos. La regresión jerárquica sobre puntuaciones compuestas reproduce la estructura del modelo, con una reducción del coeficiente del conocimiento al incorporar la autoeficacia coherente con la mediación parcial, y las covariables no alteran las relaciones focales (Tabla 8). El coeficiente focal resulta estable bajo jackknife, submuestreo, remuestreo bootstrap y winsorización, y el control negativo con un predictor placebo no produce efecto (Tabla 5-9, Figura 5-3). Esta estabilidad no atenúa las limitaciones de diseño, pero reduce la probabilidad de que la asociación dependa de observaciones influyentes o de la especificación del estimador.

5-6 CONCLUSIONES

El aporte teórico del estudio es la separación empírica, en un mismo modelo de ecuaciones estructurales, de dos vías por las que el conocimiento del razonamiento algebraico temprano se asocia con su integración en la enseñanza de las ciencias: una directa, que representa algo más de la mitad del efecto total, y otra mediada por la autoeficacia docente, que representa el 41.4 %. El aporte metodológico es un modelo de medición de tres constructos con validez verificada y una descomposición de efectos con intervalos bootstrap, transferible a otros sistemas de formación docente que integran la matemática en la enseñanza de las ciencias; el sistema peruano de formación inicial operó como contexto de contraste.

Las implicaciones para la formación inicial derivan de esa descomposición. Como ninguna vía sustituye a la otra, los programas deberían intervenir de forma simultánea sobre el conocimiento algebraico y sobre la autoeficacia para enseñarlo, por ejemplo mediante tareas de modelación de fenómenos científicos con generalización algebraica que funcionen a la vez como aprendizaje del contenido y como experiencia de dominio. Y como la ruta entre conocimiento y autoeficacia es la de mayor magnitud, la evaluación de la autoeficacia puede servir a los formadores como indicador de cuánto del conocimiento adquirido se percibe transferible a la enseñanza.

Las limitaciones acotan estas conclusiones. El diseño transversal impide establecer la dirección causal, que el modelo trata como hipótesis estructurales. Las tres variables se midieron por autoinforme en un mismo instrumento, de modo que la integración corresponde a la práctica declarada y no a la observada, y la varianza común al método no se modeló. La muestra es no probabilística por conveniencia, dos de las escalas son de construcción propia y de tres ítems, y el modelo es de un solo nivel, sin la dependencia entre participantes de una misma institución.

Quedan abiertas tres preguntas: si el peso relativo de las vías directa y mediada cambia tras las prácticas preprofesionales; qué experiencias formativas elevan la autoeficacia para enseñar el razonamiento algebraico sin limitarse a aumentar el conocimiento declarativo; y si el modelo se sostiene con medidas observadas de la práctica, muestras probabilísticas y un tratamiento multinivel de la pertenencia institucional.

5-7 REFERENCIAS BIBLIOGRÁFICAS

Aderinoye, R., Bello, A., Usman, A., Falebita, O., & Oloyede, G. (2025). Assessing the influence of primary mathematics educators' content knowledge and pedagogical content knowledge on teaching effectiveness. *Pedagogical Research, 10*(3), em0240. https://doi.org/10.29333/pr/16665

Avilés, K., & Marbán, J. (2023). Perfiles de autoeficacia docente y conocimiento especializado para la enseñanza de las matemáticas. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, *26*(2), 57-85. https://doi.org/10.6018/reifop.559321

Avilés, K., & Marbán, J. (2024). Validación de escalas de autoeficacia docente desde la perspectiva del conocimiento especializado para la enseñanza de las matemáticas. *Aula Abierta*, *53*(3), 247-256. https://doi.org/10.17811/rifie.20508

Burgos, M., Tizón, N., & Godino, J. (2024). Expanded model for elementary algebraic reasoning levels. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(7), em2475. https://doi.org/10.29333/ejmste/14753

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Gaita, R., Pino, L., & Ugarte, F. (2026). Contexts of Divisibility as an alternative to the development of Elementary Algebraic Reasoning. *Acta Scientiae*, *28*(3). https://doi.org/10.64856/acta.scientiae.8503

Gaita, R., Wilhelmi, M., Ugarte, F., & Gonzales, C. (2024). Mathematical processes for the development of algebraic reasoning in geometrical situations with in-service secondary school teachers. *Eurasia Journal*

of Mathematics, Science and Technology Education, *20*(12), em2553. https://doi.org/10.29333/ejmste/15709

Hayati, P., Mistima, M., & Sofwan, M. (2023). The Relationship between Teacher's Self-Efficacy and Creative Teaching of Primary Mathematics Teachers. *Journal of Educational and Social Research*, *13*(5), 340. https://doi.org/10.36941/jesr-2023-0141

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hermosa, J., & Sánchez, R. (2024). Avances en la caracterización del pensamiento matemático de los docentes de primaria cuando crean problemas de álgebra temprana. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, *17*(10), e12006. https://doi.org/10.55905/revconv.17n.10-389

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Ibourk, A., & Mathis, C. (2024). Developing preservice elementary teachers' self-efficacy toward teaching science. *International Journal of Science Education*, *47*(5), 656-679. https://doi.org/10.1080/09500693.2024.2347154

Jerrim, J., Prieto, C., Marcenaro, O., & Shure, N. (2024). Teacher Self-Efficacy, Instructional Practice, and Student Outcomes: Evidence

from the TALIS Video Study. *American Educational Research Journal*, *62*(2), 378-413. https://doi.org/10.3102/00028312241300265

Lee, M., & Lee, J. (2023). The relationship between preservice elementary teachers' solutions to a pattern generalization problem and difficulties they anticipate in teaching it. *The Journal of Mathematical Behavior*, *69*, 101046. https://doi.org/10.1016/j.jmathb.2023.101046

Leijen, Ä., Pedaste, M., Baucal, A., Poom, K., & Lepp, L. (2024). What predicts instructional quality and commitments to teaching: self-efficacy, pedagogical knowledge or integration of the two?. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1287313

Lumbantoruan, J. (2023). The Influence of Opportunities to Learn and Efficacy Belief Factor towards Mathematical Knowledge for Teaching. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, *7*(3), 676. https://doi.org/10.31764/jtam.v7i3.15013

Lupión, T., García, C., & Toma, R. (2026). Preservice Primary Teachers' Attitudes Towards STEM Education: a Study Developed Through an Integrated STEM Training Programme Using Inquiry- Based Approaches and Engineering Design Practices. *Research in Science Education*, *56*(5), 1443-1466. https://doi.org/10.1007/s11165-026-10323-y

López, Y., Navarro, B., López, C., & Chambergo, A. (2026). Percepción de la matemática y rendimiento académico en la formación inicial del docente de primaria. *Revista Simón Rodríguez*, *6*(11), 458-470. https://doi.org/10.62319/simonrodriguez.v.6i11.143

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Melissa, M., Anuniwat, D., & Pratini, H. (2026). A review of mathematics teachers' noticing of students' algebraic thinking. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1897747

Menon, D., Shorman, D., Cox, D., & Thomas, A. (2023). Preservice Elementary Teachers Conceptions and Self-Efficacy for Integrated STEM. *Education Sciences, 13*(5), 529. https://doi.org/10.3390/educsci13050529

Menon, D., Wieselmann, J., Haines, S., Asim, S., & Johnson, A. (2026). "I learned a lot, but…": preservice elementary teachers' integrated STEM teaching self-efficacy and perceived challenges. *International Journal of STEM Education, 13*(1). https://doi.org/10.1186/s40594-026-00645-8

Menon, D., Wieselmann, J., Haines, S., Asim, S., Koch, A., & Cox, D. (2025). Preservice Elementary Teachers' Integrated STEM Teaching Self-Efficacy: Contributing Sources Within STEM Education Courses. *AERA Open, 11*. https://doi.org/10.1177/23328584251321472

Narváez, R., Adamuz, N., & Cañadas, M. (2025). Análisis bibliométrico sobre pensamiento algebraico en educación infantil y primaria en Scopus. *Avances de Investigación en Educación Matemática*(27), 43-65. https://doi.org/10.35763/aiem27.5825

Özaltun, A., & Bukova, E. (2026). Preservice mathematics teachers' confidence in teaching algebra: A scale adaptation study. *Hungarian Educational Research Journal, 16*(2), 260-274. https://doi.org/10.1556/063.2026.00479

Öztürk, N., & Işler, I. (2024). Development of prospective elementary teachers' knowledge to teach early algebra through case discussions. *Journal of Mathematics Teacher Education, 28*(6), 1341-1372. https://doi.org/10.1007/s10857-024-09642-6

Pallauta, J., Batanero, C., Rivas, M., & Vera, O. (2026). Competencia de las futuras personas docentes en resolución, creación y razonamiento algebraico de problemas sobre tablas estadísticas. *Uniciencia, 40*(1), 1-23. https://doi.org/10.15359/ru.40-1.2

Pallauta, J., Rivas, M., Batanero, C., & Gea, M. (2025). Razonamiento Algebraico de Profesores en Formación. *Journal of Research in Mathematics Education,* 1-22. https://doi.org/10.17583/redimat.16710

Pincheira, N., & Alsina, Á. (2024). Mathematical Knowledge of Early Algebra Exhibited by Pre-Service Early Childhood Education Teachers. *International Journal of Science and Mathematics Education, 23*(2), 461-487. https://doi.org/10.1007/s10763-024-10478-y

Pincheira, N., & Alsina, Á. (2026). Knowledge for teaching early algebra in primary education: Integral analysis from the MKT. *International Electronic Journal of Mathematics Education, 21*(2), em0881. https://doi.org/10.29333/iejme/18554

Pinto, E., & Cortés, C. (2025). Club de vídeos para mirar profesionalmente el pensamiento algebraico de estudiantes. *Enseñanza de las Ciencias.*

Revista de investigación y experiencias didácticas, 43(3), 201-219. https://doi.org/10.5565/rev/ensciencias.6243

Pinto, E., Ayala, C., Molina, M., & Cañadas, M. (2023). Desarrollo del pensamiento algebraico a través de la justificación en educación primaria. *Enseñanza de las Ciencias. Revista de investigación y experiencias didácticas, 41*(1), 149-173. https://doi.org/10.5565/rev/ensciencias.5835

Razali, H., Jamaluddin, R., & Kamarudin, N. (2022). Implementing Integrated STEM Teaching in Design and Technology: Teachers' Knowledge and Teaching Practices. *International Journal of Academic Research in Business and Social Sciences, 12*(9). https://doi.org/10.6007/ijarbss/v12-i9/14785

Ribeiro, A., Aguiar, M., Trevisan, A., & Elias, H. (2024). Exploring Learning Opportunities for Primary Teachers: the Case of Knowledge for Teaching Early Algebra. *Revista Latinoamericana de Investigación en Matemática Educativa, 26*(3), 327-356. https://doi.org/10.12802/relime.23.2633

Ricaldi, M. (2026). Percepciones de profesores de matemáticas en el Perú sobre su formación profesional. *Revista Colombiana de Educación*(101), e20445. https://doi.org/10.17227/rce.num101-20445

Shahali, E., & Halim, L. (2023). The Influence of Science Teachers' Beliefs, Attitudes, Self-efficacy and School Context on Integrated STEM Teaching Practices. *International Journal of Science and Mathematics Education, 22*(4), 787-807. https://doi.org/10.1007/s10763-023-10403-9

Soler, M., & Valencia, N. (2025). Autoeficacia de profesores de matemáticas en formación. Revisión de la literatura en Latinoamérica. *Tecné, Episteme y Didaxis: TED*(58). https://doi.org/10.17227/ted.num58-22975

Sun, S., Sun, D., & Xu, T. (2023). The Developmental Progression of Early Algebraic Thinking of Elementary School Students. *Journal of Intelligence*, *11*(12), 222. https://doi.org/10.3390/jintelligence11120222

Velez, M., Mamani, O., Zerga, C., Turpo, J., & Morales, W. (2024). Effect of perceived stress, job satisfaction, and workload on the professional self-efficacy of Peruvian regular basic education teachers. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1302624

CAPÍTULO VI

PRESIÓN ACADÉMICA, OPORTUNIDAD PERCIBIDA Y CLARIDAD DE LA POLÍTICA DE IA COMO DETERMINANTES DE LA INTENCIÓN DE USO INDEBIDO DE LA IA GENERATIVA EN LA EDUCACIÓN SUPERIOR: EL PAPEL MEDIADOR DE LA DESCONEXIÓN MORAL

Mauricio Acevedo Carrillo

Universidad Privada del Norte

Resumen

La inteligencia artificial generativa (IAG) ha abaratado la externalización de trabajos evaluables y ha dificultado su detección. Este estudio integra el triángulo del fraude con la teoría sociocognitiva de la agencia moral para explicar la intención de uso indebido de la IAG y pone a prueba si la desconexión moral es el mecanismo mediante el cual la presión académica, la oportunidad percibida y la claridad de la política institucional de IA configuran dicha intención. Se realizó una encuesta transversal en línea con 310 estudiantes de pregrado de universidades públicas y privadas de Perú (155 de áreas STEM y 155 de áreas no STEM), entre abril y junio de 2026, con tamaño muestral calculado a priori. El modelo, de cinco constructos reflectivos, se estimó mediante PLS-SEM con 5000 remuestreos bootstrap, PLSpredict y análisis multigrupo con invarianza de medida de los compuestos. El modelo de medida cumplió los criterios de fiabilidad y validez. La desconexión moral fue el predictor más fuerte de la intención ($\beta = .352$); la presión y la oportunidad la incrementaron y la claridad de la política la redujo ($\beta = -.294$). Esta última carecía de efecto directo significativo sobre la intención y su efecto total (−.176) se transmitió por la desconexión moral. El modelo explicó el 37.1 % de la varianza de la intención, predijo mejor que un modelo lineal de referencia y no difirió entre grupos disciplinares. Los resultados sugieren que las normas claras, formuladas a nivel de cada evaluación, protegen la integridad al reducir las justificaciones morales disponibles.

Palabras clave: integridad académica; inteligencia artificial generativa; desconexión moral; triángulo del fraude; política de educación superior; PLS-SEM.

Abstract

Generative artificial intelligence (GenAI) has made the undisclosed outsourcing of assessed work cheap and hard to detect. This study integrates the fraud triangle with the social-cognitive theory of moral agency to explain students' intention to

misuse GenAI, and tests whether moral disengagement is the mechanism through which academic pressure, perceived opportunity and institutional AI-policy clarity shape that intention. An online cross-sectional survey was conducted between April and June 2026 with 310 undergraduates from public and private universities in Peru (155 STEM and 155 non-STEM), with a sample size calculated a priori. The model, comprising five reflective constructs, was estimated with PLS-SEM using 5,000 bootstrap resamples, PLSpredict and a multigroup analysis with measurement invariance of composites. The measurement model met all reliability and validity criteria. Moral disengagement was the strongest predictor of intention (β = .352); pressure and opportunity increased it, whereas policy clarity reduced it (β = −.294). Policy clarity had no significant direct effect on intention, and its total effect (−.176) was transmitted through moral disengagement. The model explained 37.1% of the variance in intention, predicted better than a linear benchmark and did not differ across disciplinary groups. The findings suggest that clear rules, formulated at the level of each assessment, protect integrity by reducing the moral justifications available to students.

Keywords: *academic integrity; generative artificial intelligence; moral disengagement; fraud triangle; higher education policy; PLS-SEM.*

6.1 INTRODUCCIÓN

La inteligencia artificial generativa (IAG) ha modificado la economía de la conducta académica indebida. Redactar un ensayo, un informe o las respuestas de un cuestionario en línea ya no exige recurrir a un servicio de compra de trabajos ni a un compañero dispuesto a compartir el suyo: un modelo conversacional produce un texto aceptable en segundos, sin coste y con un estilo que los detectores de similitud no fueron diseñados para señalar (Cotton et al., 2024; Eke, 2023). Las universidades han encuadrado la IAG como una amenaza para la integridad académica, aunque reconocen su valor legítimo para el aprendizaje (Yusuf et al., 2024; Song, 2024). El estudiantado informa de un uso frecuente de agentes conversacionales en tareas evaluables y de incertidumbre sobre dónde termina la ayuda y empieza la falta (Gruenhagen et al., 2024; Fajt y Schiller,

2025); con frecuencia, la herramienta se emplea para aligerar las cargas del aprendizaje y no para potenciarlo (Yeung et al., 2026), y las evaluaciones auténticas, antes consideradas resistentes a la externalización, también son vulnerables (Kofinas et al., 2025). Un estudio de seguimiento en educación secundaria halló que las tasas globales de trampa se mantuvieron estables, lo que sugiere que la disponibilidad de la IAG modifica los medios más que la prevalencia (Chen et al., 2026), y la investigación sobre el uso excesivo apunta a que las motivaciones del estudiantado condicionan sus consecuencias (Abbas et al., 2025).

La respuesta institucional ha adoptado sobre todo la forma de políticas y directrices. Los análisis de documentos institucionales muestran enfoques heterogéneos, desde la prohibición hasta la integración regulada, y una ambigüedad frecuente sobre lo que se espera del estudiante en cada evaluación (Jin et al., 2025; An et al., 2025). El examen de 142 políticas de integridad reveló un vacío notable en la mención de la IA y de las tecnologías asociadas (Perkins y Roe, 2024), y se ha señalado que la mayoría de las instituciones carece de directrices formales, lo que genera confusión (Song, 2024). Lo que permanece sin aclarar es cómo, en el nivel del estudiante individual, la claridad de estas normas se traduce en conducta.

La investigación sobre los determinantes individuales del uso indebido de la IAG crece con rapidez, pero sigue fragmentada (Yilmaz, 2026). Los estudios basados en el triángulo del fraude indican que la presión, la oportunidad y la racionalización se asocian con el mal uso de ChatGPT (Alshurafat et al., 2024; Rahman et al., 2026); los modelos basados en la teoría del comportamiento planificado destacan la intención como resultado proximal de la deshonestidad y del plagio (Al-Nuaimi y Uzun, 2023; Duong et al., 2026); y los modelos de percepción destacan el riesgo percibido y las actitudes (Ortiz y Blahopoulou, 2025; Ofem et al., 2025), así como el hábito en la aceptación de la herramienta (Strzelecki, 2024). Tres limitaciones recorren esta literatura. Primera, la racionalización suele entrar como un predictor paralelo, aunque la teoría sociocognitiva describe las autojustificaciones que desactivan las autosanciones morales

como el *mecanismo* por el que las fuerzas situacionales se convierten en conducta indebida (Bandura et al., 1996; Ogunfowora et al., 2022), y la evidencia sitúa la desconexión moral y las técnicas de neutralización entre los correlatos más fuertes de la deshonestidad académica (Sirca y Billen, 2024; Flutti et al., 2026). Segunda, la política institucional se mide como existencia o conocimiento y su efecto es inconsistente, lo que deja abierto si la claridad de las normas actúa directamente sobre la intención o a través de las justificaciones disponibles. Tercera, la mayoría de los modelos se evalúa solo dentro de la muestra, y las diferencias disciplinares se suponen en lugar de contrastarse mediante invarianza.

Este estudio aborda dichas limitaciones con un modelo basado en el mecanismo, en el que la desconexión moral media los efectos de la presión académica, la oportunidad percibida y la claridad de la política de IA sobre la intención de uso indebido de la IAG. El modelo se estima con PLS-SEM, se evalúa su capacidad predictiva fuera de muestra y se contrasta su invarianza entre estudiantes de áreas STEM y no STEM. Perú actúa como escenario de prueba: las universidades latinoamericanas están infrarrepresentadas en la evidencia revisada, que procede sobre todo de Europa, Asia, África, Oceanía y Norteamérica, y los trabajos regionales apenas comienzan a abordar la IAG y la integridad, por ejemplo mediante el análisis de la regulación sudamericana, en la que no se identificó una norma específica para garantizar la integridad académica (Cabrera et al., 2026). La contribución es, por tanto, teórica y metodológica más que regional: se especifica y contrasta la vía cognitiva por la que las condiciones situacionales e institucionales se convierten en intenciones de conducta indebida, con la evidencia predictiva y de invarianza de la que el campo carecía.

6.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

6.2.1 El triángulo del fraude y la desconexión moral

El triángulo del fraude explica la conducta deshonesta como resultado conjunto de presión, oportunidad y racionalización, y se ha trasladado del fraude ocupacional a la deshonestidad académica y al uso indebido de la

IAG (Alshurafat et al., 2024; Rahman et al., 2026). La teoría sociocognitiva refina su tercer elemento: la desconexión moral reúne mecanismos cognitivos, como la justificación moral, la comparación ventajosa, el desplazamiento de la responsabilidad y la distorsión de las consecuencias, que desactivan la autorregulación moral y permiten actuar contra los propios estándares sin autocensura (Bandura et al., 1996). La evidencia metaanalítica la asocia de forma consistente con la conducta indebida y la sitúa como vía de transmisión de sus antecedentes (Ogunfowora et al., 2022). La variable resultado es la intención de uso indebido, antecedente proximal de la acción.

6.2.2 Presión académica

La presión remite a las exigencias percibidas de calificaciones, carga de trabajo y consecuencias del fracaso. Es un determinante significativo del mal uso de ChatGPT en los modelos del triángulo del fraude (Alshurafat et al., 2024; Rahman et al., 2026) y ofrece justificaciones preparadas, como una carga injusta, que hacen defendible la falta.

- *H1a.* La presión académica se relaciona positivamente con la desconexión moral.
- *H1b.* La presión académica se relaciona positivamente con la intención de uso indebido de la IAG.

6.2.3 Oportunidad percibida

La oportunidad es la facilidad percibida para usar indebidamente la IAG sin ser detectado ni sancionado. La dificultad de detectar el texto generado por IA es una preocupación recurrente (Cotton et al., 2024; Eke, 2023), y el estudiantado percibe los agentes conversacionales como fácilmente utilizables en las evaluaciones (Gruenhagen et al., 2024). Un riesgo de detección bajo favorece, además, las comparaciones ventajosas y la distorsión de las consecuencias.

- *H2a.* La oportunidad percibida se relaciona positivamente con la desconexión moral.
- *H2b.* La oportunidad percibida se relaciona positivamente con la intención de uso indebido de la IAG.

6.2.4 Claridad de la política de IA

La claridad de la política es el grado en que el estudiantado conoce los usos permitidos de la IAG, cómo declararlos y las consecuencias de su abuso. Las normas claras eliminan la ambigüedad que explota el desplazamiento de la responsabilidad, y los análisis institucionales identifican la claridad y la comunicación a nivel de tarea como el punto débil de las políticas actuales (Jin et al., 2025; An et al., 2025; Perkins y Roe, 2024). Dado que el efecto directo de las políticas ha sido inconsistente (Rahman et al., 2026), el modelo especifica un camino directo y otro indirecto.

- *H3a.* La claridad de la política de IA se relaciona negativamente con la desconexión moral.
- *H3b.* La claridad de la política de IA se relaciona negativamente con la intención de uso indebido de la IAG.

6.2.5 La desconexión moral como mecanismo

Si la desconexión moral es el mecanismo por el que las fuerzas situacionales se convierten en conducta indebida, debe predecir la intención y transmitir los efectos de los tres antecedentes.

- *H4.* La desconexión moral se relaciona positivamente con la intención de uso indebido de la IAG.
- *H5a–c.* La desconexión moral media las relaciones de la presión académica (H5a), la oportunidad percibida (H5b) y la claridad de la política de IA (H5c) con la intención.

Como el uso habitual rebaja el umbral de uso en situaciones nuevas y se asocia con la aceptación y el uso problemático de la IAG (Strzelecki, 2024; Abbas et al., 2025), la frecuencia de uso se incluye como control.

6.2.6 Heterogeneidad disciplinar

Las disciplinas difieren en formatos de evaluación y en la facilidad para presentar como propio el producto de la IAG, y los estudios multigrupo sobre ChatGPT y deshonestidad informan diferencias en algunas relaciones (Ofem et al., 2025). Sin formular una hipótesis direccional, se plantea la pregunta de investigación *PI1*: ¿difieren las relaciones estructurales del modelo entre estudiantes de áreas STEM y no STEM?

6.3 MATERIALES Y MÉTODOS

6.3.1 Diseño, participantes y tamaño muestral

Se realizó una encuesta transversal explicativo-predictiva en línea entre abril y junio de 2026 con estudiantes de pregrado de 18 años o más, de universidades públicas y privadas de Perú. Mediante cuotas no probabilísticas se obtuvo igual número de estudiantes STEM (ingeniería, ciencias de la salud, ciencias naturales y exactas, computación) y no STEM (economía y empresa, educación, derecho y ciencia política, ciencias sociales y humanidades), con invitaciones por listas institucionales y anuncios de curso; el diseño balanceado por cuotas disciplinares tiene precedentes en la región (Torres et al., 2026). Al no existir selección probabilística, no se infieren proporciones poblacionales.

El tamaño muestral se fijó antes de la recogida para cada análisis planificado, adoptando el mayor requisito (Tabla 6-1). Para los caminos estructurales se aplicó el método de la raíz cuadrada inversa (Kock y Hadaya, 2018) con α unilateral = .05, potencia = .80 y β mínimo = .15, valor que corresponde a la relación más débil esperada, la de la política institucional, no significativa en el modelo más reciente del triángulo del fraude (Rahman et al., 2026). El R^2 anticipado de la intención fue .43, límite inferior de los modelos

comparables. El análisis multigrupo, con $\beta = .20$ por grupo, resultó el más exigente y fijó N = 310 (155 por grupo). El efecto indirecto específico más pequeño se evaluó mediante simulación de Monte Carlo de la prueba de significación conjunta (Schoemann et al., 2017).

Tabla 6-1

Cálculo a priori del tamaño muestral para los análisis del estudio

Análisis	Criterio y Parámetros	N mínimo	Potencia con N = 310
Camino estructural más pequeño (modelo agregado)	Raíz cuadrada inversa: $N \geq ((z_{0.95} + z_{0.80})/\beta_min)^2 = (1.64485 + 0.84162)^2/0.15^2 = 274.78$; α unilateral = .05	275	.840
Caminos focales dentro de cada grupo disciplinar (AMG)	Raíz cuadrada inversa por grupo con $\beta_min = 0.20$: $n_g \geq (2.48647/0.20)^2 = 154.56$; dos grupos de igual tamaño	310 (155 por grupo)	.801
Coeficiente único en la ecuación de INT	Prueba F, F no central; $f^2 = \beta^2/(1 - R^2) = 0.15^2/(1 - 0.43) = 0.0395$; u = 1; k = 5 predictores; $\alpha = .05$; potencia = .80	201	.937
Varianza explicada de INT	Prueba F, F no central; $f^2 = 0.15$; u = 5; k = 5; $\alpha = .05$; potencia = .80	92	1.000
Varianza explicada de MD	Prueba F, F no central; $f^2 = 0.15$; u = 3; k = 3; $\alpha = .05$; potencia = .80	77	1.000
Efecto indirecto específico más pequeño (POL → MD → INT)	Prueba de significación conjunta por Monte Carlo; a = −0.20, b = 0.40, α bilateral = .05, 1000 réplicas por N candidato	195	.950

Nota: El tamaño muestral del estudio es el mayor de los mínimos, N = 310, fijado por el análisis multigrupo. La potencia con N = 310 de la fila de Monte Carlo se estimó con 4000 réplicas. β_min = camino estandarizado más pequeño de interés sustantivo; R^2 = varianza explicada anticipada de INT; u = grados de libertad del numerador; k = número de predictores; AMG = análisis multigrupo; MD = desconexión moral; INT = intención de uso indebido; POL = claridad de la política de IA.

6.3.2 Medidas y procedimiento

Los cinco constructos fueron reflectivos, con cuatro indicadores en escala de acuerdo de siete puntos (Tabla 6-3). La presión académica (PRE) y la oportunidad percibida (OPP) se adaptaron de operacionalizaciones del triángulo del fraude; la claridad de la política (POL) cubrió usos permitidos, explicación por tarea, declaración y consecuencias; la desconexión moral

(MD) incluyó un ítem por mecanismo adaptado al uso de IAG en tareas, a partir de una escala de desconexión moral asociada al plagio validada con universitarios peruanos (Lingán et al., 2026); y la intención de uso indebido (INT) siguió la formulación de la intención conductual (Ajzen, 1991; Pérez et al., 2026). La frecuencia de uso (USE; 1 = nunca, 5 = a diario) actuó como control. El cuestionario, administrado en español, fue anónimo y de participación voluntaria; los bloques de antecedentes, mediador y resultado se presentaron en pantallas separadas y los ítems de intención se redactaron como conducta hipotética. Como control estadístico del sesgo de método común se aplicó la prueba de colinealidad completa, con VIF > 3.3 como señal de alerta (Kock, 2015).

6.3.3 Análisis estadístico

Se empleó PLS-SEM con modo A, esquema de ponderación de caminos, indicadores estandarizados y criterio de convergencia de 10^{-7}, dado que el objetivo combina explicación y predicción (Hair et al., 2021). La inferencia utilizó 5000 remuestreos bootstrap: intervalos percentiles para cargas, HTMT y términos de robustez, e intervalos corregidos por sesgo y acelerados (BCa) para caminos y efectos directos, indirectos y totales. El modelo de medida se evaluó con cargas (≥ .708), α de Cronbach, ρA y ρC (.70–.95), AVE (≥ .50), HTMT (< .85; Henseler et al., 2015) y criterio de Fornell–Larcker. El modelo estructural se evaluó con VIF interno (< 3), caminos, f^2, R^2 y PLSpredict (10 pliegues × 10 repeticiones) frente a un modelo lineal de referencia (Hair et al., 2021); el ajuste global, con el SRMR (< .08; Schuberth et al., 2023). La mediación se clasificó como complementaria, competitiva o solo indirecta, con la varianza explicada (VAF) como razón entre el efecto indirecto y el total.

La robustez se examinó añadiendo, uno a uno, términos cuadráticos en dos etapas y contrastando la endogeneidad con cópulas gaussianas, identificadas solo para regresores no normales según la prueba de Lilliefors (Becker et al., 2022; Sarstedt et al., 2022). La heterogeneidad disciplinar se analizó con el procedimiento de invarianza de medida de los compuestos en tres pasos (MICOM) y un análisis multigrupo por permutación (5000

permutaciones), con valores p ajustados por Holm en 11 contrastes (Henseler et al., 2016). Los cálculos se realizaron con rutinas propias en Python 3 con semilla fija.

6.4 RESULTADOS

6.4.1 Participantes

Tabla 6-2

Perfil de los participantes por grupo disciplinar

Característica	Categoría	STEM (n = 155)	No STEM (n = 155)	Total (N = 310)
Género	Mujer	80 (51.6)	89 (57.4)	169 (54.5)
	Hombre	72 (46.5)	62 (40.0)	134 (43.2)
	No binario o prefiere no indicarlo	3 (1.9)	4 (2.6)	7 (2.3)
Edad (años)	M (DE); Mdn	21.0 (3.0); 20	21.2 (3.0); 20	21.1 (3.0); 20
Universidad	Pública	94 (60.6)	82 (52.9)	176 (56.8)
	Privada	61 (39.4)	73 (47.1)	134 (43.2)
Año de estudios	1	30 (19.4)	29 (18.7)	59 (19.0)
	2	40 (25.8)	35 (22.6)	75 (24.2)
	3	34 (21.9)	39 (25.2)	73 (23.5)
	4	31 (20.0)	33 (21.3)	64 (20.6)
	5	20 (12.9)	19 (12.3)	39 (12.6)
Área	Ingeniería	70 (45.2)	0 (0.0)	70 (22.6)
	Ciencias de la salud	32 (20.6)	0 (0.0)	32 (10.3)
	Ciencias naturales y exactas	21 (13.5)	0 (0.0)	21 (6.8)
	Computación y sistemas de información	32 (20.6)	0 (0.0)	32 (10.3)
	Economía y empresa	0 (0.0)	58 (37.4)	58 (18.7)
	Educación	0 (0.0)	32 (20.6)	32 (10.3)
	Derecho y ciencia política	0 (0.0)	25 (16.1)	25 (8.1)
	Ciencias sociales y humanidades	0 (0.0)	40 (25.8)	40 (12.9)

Característica	Categoría	STEM (n = 155)	No STEM (n = 155)	Total (N = 310)
Uso de IAG en tareas académicas	Nunca	15 (9.7)	13 (8.4)	28 (9.0)
	Menos de una vez al mes	28 (18.1)	34 (21.9)	62 (20.0)
	Mensual	46 (29.7)	52 (33.5)	98 (31.6)
	Semanal	50 (32.3)	49 (31.6)	99 (31.9)
	Diario	16 (10.3)	7 (4.5)	23 (7.4)

Nota: Los valores son n (%) salvo indicación contraria. STEM = ciencias, tecnología, ingeniería, matemáticas y ciencias de la salud; IAG = inteligencia artificial generativa.

La Tabla 6-2 describe a los 310 participantes. Las mujeres representaron el 54.5 % de la muestra, la edad media fue de 21.1 años (DE = 3.0) y el 56.8 % cursaba estudios en una universidad pública. El 31.9 % y el 7.4 % declararon usar la IAG semanal y diariamente en tareas académicas, respectivamente, y el 9.0 % nunca la usaba.

6.4.2 Modelo de medida

Todas las cargas externas se situaron entre .784 y .884, por encima de .708, con límites inferiores bootstrap superiores a .70, y la colinealidad de los indicadores fue baja (VIF máximo = 2.52; Tabla 6-3).

Tabla 6-3

Redacción de los indicadores, estadísticos descriptivos, cargas externas y colinealidad

Constructo	Ítem	Redacción	M	DE	Carga [IC 95 %]	VIF
PRE	PRE1	Siento una gran presión por obtener buenas calificaciones este semestre	4.91	1.29	.850 [.803, .885]	2.30
	PRE2	Las fechas límite de mis evaluaciones suelen superponerse y me dejan poco tiempo.	4.90	1.22	.844 [.800, .878]	1.98
	PRE3	Suspender una asignatura tendría graves consecuencias para mí.	4.89	1.29	.874 [.836, .902]	2.43
	PRE4	La carga de trabajo de mi carrera supera mi capacidad.	4.94	1.28	.805 [.751, .848]	1.75

Constructo	Ítem	Redacción	M	DE	Carga [IC 95 %]	VIF
OPP	OPP1	Es poco probable que mis profesores se den cuenta del texto generado por IA en una tarea.	4.35	1.29	.813 [.758, .856]	1.82
	OPP2	La mayoría de las evaluaciones de mi programa se pueden completar con IA generativa sin que nadie se dé cuenta.	4.30	1.33	.858 [.821, .888]	2.12
	OPP3	Las herramientas de detección que usa mi universidad se pueden eludir fácilmente.	4.31	1.26	.842 [.799, .877]	1.95
	OPP4	Las probabilidades de ser sancionado por usar IA sin declarar son bajas.	4.27	1.35	.784 [.718, .833]	1.67
POL	POL1	Mi universidad tiene normas claras sobre qué usos de la IA generativa están permitidos en los trabajos evaluados.	3.57	1.32	.884 [.852, .909]	2.52
	POL2	Mis profesores explican cómo se puede usar la IA generativa en cada trabajo.	3.57	1.33	.821 [.762, .863]	1.89
	POL3	Sé cómo informar sobre el uso de IA generativa en mis trabajos.	3.65	1.34	.849 [.800, .884]	2.27
	POL4	Las consecuencias del mal uso de la IA generativa se comunican claramente.	3.54	1.37	.815 [.757, .857]	1.82
MD	MD1	Usar IA para redactar parte de una tarea es aceptable cuando la carga de trabajo es excesiva.	3.14	1.29	.848 [.809, .878]	2.13
	MD2	Presentar textos generados por IA no es peor que lo que hacen otros estudiantes.	3.10	1.29	.879 [.851, .903]	2.45
	MD3	Si la universidad no lo prohíbe explícitamente, no se puede culpar a los estudiantes por usar IA.	3.14	1.26	.817 [.772, .852]	1.91
	MD4	El uso no divulgado de IA en trabajos académicos no perjudica a nadie.	3.19	1.26	.847 [.810, .878]	2.12

Constructo	Ítem	Redacción	M	DE	Carga [IC 95 %]	VIF
INT	INT1	El próximo ciclo, tengo la intención de presentar textos generados por IA como trabajo propio si fuera necesario.	2.97	1.25	.876 [.848, .897]	2.30
	INT2	Utilizaría IA generativa para completar una evaluación en la que su uso no esté permitido.	2.97	1.27	.874 [.847, .898]	2.45
	INT3	Planeo usar IA generativa para responder exámenes o cuestionarios en línea.	2.99	1.26	.839 [.797, .873]	2.16
	INT4	Si tuviera poco tiempo, dejaría que la IA escribiera la mayor parte de una tarea sin revelarlo.	2.99	1.26	.827 [.781, .864]	2.02

Nota: Escala de acuerdo de siete puntos (1 = totalmente en desacuerdo; 7 = totalmente de acuerdo); los ítems se muestran en la redacción de trabajo en inglés del cuestionario administrado en español. Los intervalos de confianza son intervalos percentiles bootstrap de 5000 remuestreos. VIF = factor de inflación de la varianza del indicador dentro de su bloque. USE se midió con un único ítem de frecuencia de cinco puntos (1 = nunca; 5 = a diario). PRE = presión académica; OPP = oportunidad percibida; POL = claridad de la política de IA; MD = desconexión moral; INT = intención de uso indebido.

La fiabilidad de consistencia interna y la validez convergente se situaron dentro de lo recomendado (α de Cronbach de .843 a .877; AVE de .680 a .730), y los VIF de colinealidad completa no superaron 1.59, por debajo de 3.3 (Tabla 6-4).

Tabla 6-4

Fiabilidad de consistencia interna, validez convergente y colinealidad completa

Constructo	M	DE	α de Cronbach	ρA	ρC	AVE	VIF de colinealidad completa
Presión académica (PRE)	4.91	1.07	.865	.868	.908	.712	1.19
Oportunidad percibida (OPP)	4.31	1.08	.843	.846	.895	.680	1.33
Claridad de la política de IA (POL)	3.58	1.13	.864	.868	.907	.710	1.24
Desconexión moral (MD)	3.14	1.08	.870	.871	.911	.719	1.52
Intención de uso indebido (INT)	2.98	1.08	.877	.886	.915	.730	1.59
Frecuencia de uso de IAG (USE)	3.09	1.08	—	—	—	—	1.23

Nota: M y DE se calcularon sobre la media de los indicadores de cada constructo (escala 1–7; USE, 1–5). ρA = fiabilidad de Dijkstra–Henseler; ρC = fiabilidad compuesta; AVE = varianza media extraída. VIF de colinealidad completa = factor de inflación de la varianza de la puntuación de cada constructo regresada sobre las demás.

La razón HTMT más alta correspondió al par MD–INT (.577; límite superior .667), muy por debajo de .85, y la raíz cuadrada de cada AVE superó las correlaciones del constructo con los demás (Tabla 6-5).

Tabla 6-5

Validez discriminante: razones heterorrasgo-monorrasgo y criterio de Fornell–Larcker

Constructo	PRE	OPP	POL	MD	INT
Presión académica (PRE)	**.844**	.128 [.267]	.203 [.321]	.386 [.505]	.390 [.497]
Oportunidad percibida (OPP)	.109	**.825**	.321 [.441]	.299 [.421]	.431 [.544]
Claridad de la política de IA (POL)	−.176	−.276	**.843**	.440 [.556]	.344 [.456]
Desconexión moral (MD)	.335	.257	−.382	**.848**	.577 [.667]
Intención de uso indebido (INT)	.343	.373	−.301	.511	**.854**

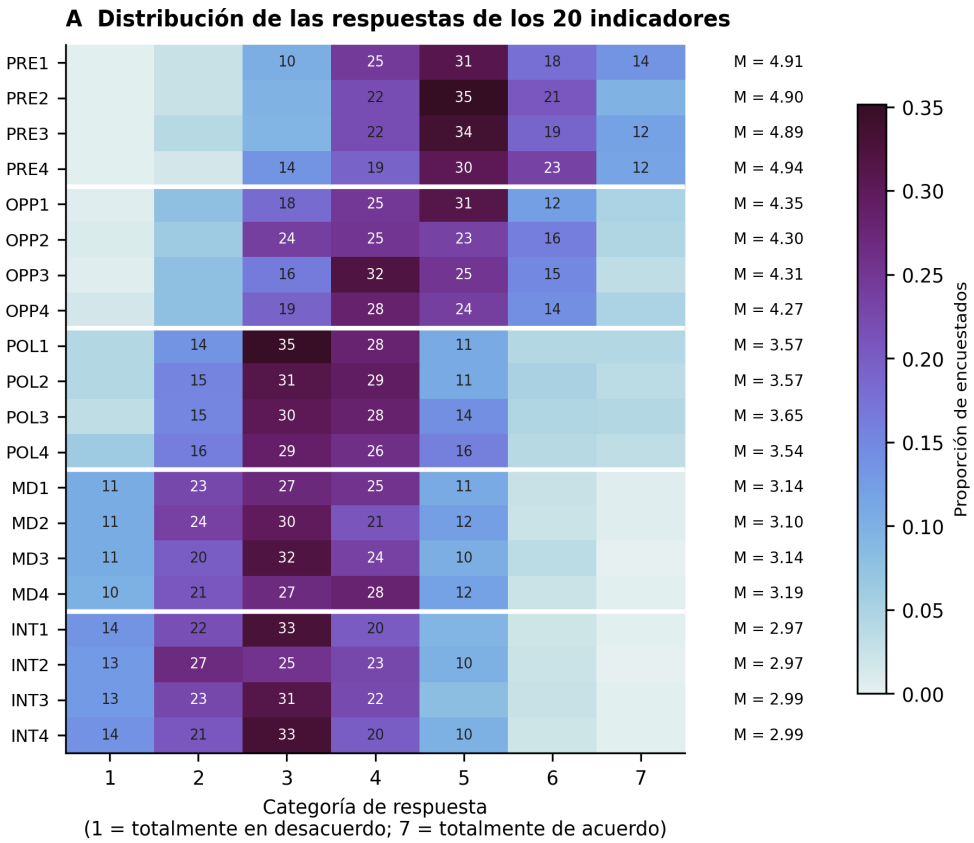
Nota: Diagonal (negrita): raíz cuadrada de la AVE. Bajo la diagonal: correlaciones entre puntuaciones de los constructos. Sobre la diagonal: razón HTMT con el límite superior de su intervalo percentil bootstrap del 95 % entre corchetes.

6.4.3 Modelo estructural

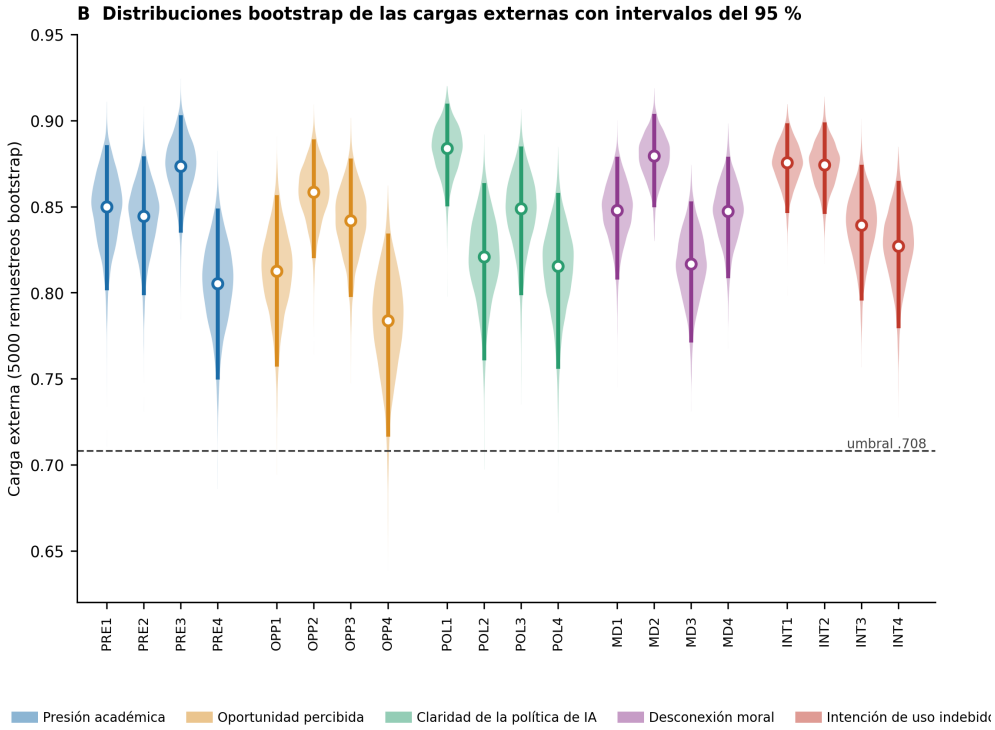
La colinealidad interna fue baja (VIF máximo = 1.32). La desconexión moral fue el predictor más fuerte de la intención, con apoyo a H4 (Tabla 6-6). La presión y la oportunidad se relacionaron positivamente con la desconexión moral y la claridad de la política lo hizo negativamente, con apoyo a H1a, H2a y H3a. La presión y la oportunidad mostraron además relaciones directas positivas con la intención (H1b y H2b apoyadas), mientras que la relación directa de la claridad de la política no fue significativa, por lo que H3b no recibió apoyo. La frecuencia de uso se relacionó positivamente con la intención. Todos los tamaños del efecto fueron pequeños; los mayores correspondieron a MD → INT (f^2 = .149), próximo al umbral de efecto mediano, y a POL → MD (f^2 = .103).

Figura 6-1

Modelo de medida de la presión académica, la oportunidad percibida, la claridad de la política de IA, la desconexión moral y la intención de uso indebido.

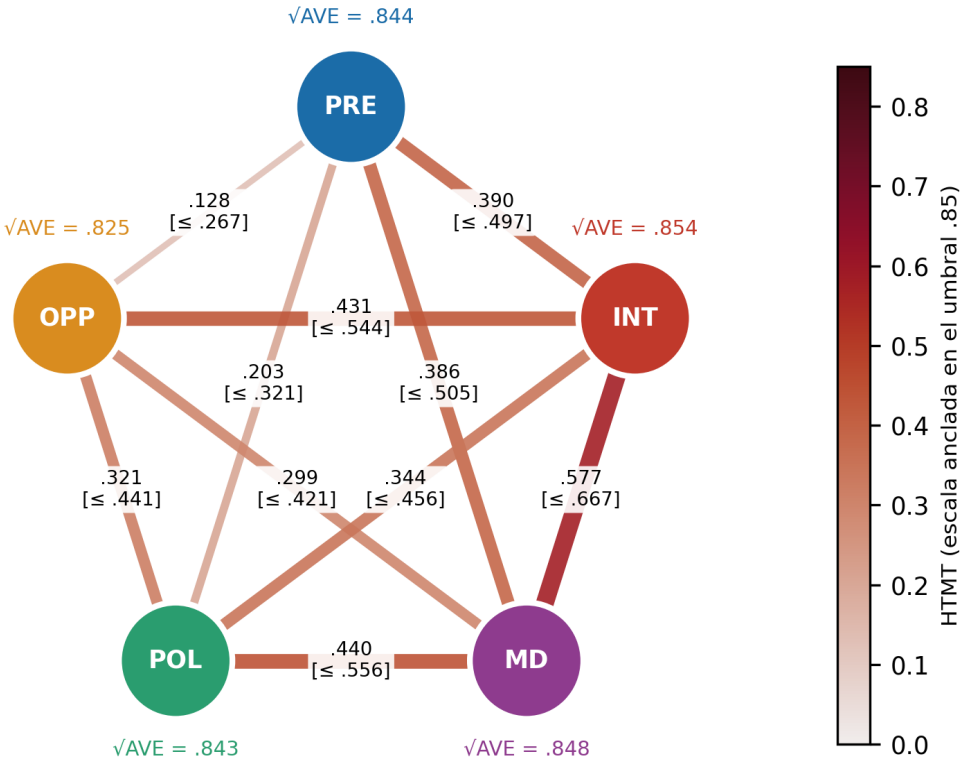


Nota. N = 310. Panel A: proporción de encuestados en cada categoría de la escala de siete puntos para los 20 indicadores (se rotulan con el porcentaje las celdas con valores de al menos el 10 %), con la media del indicador a la derecha.

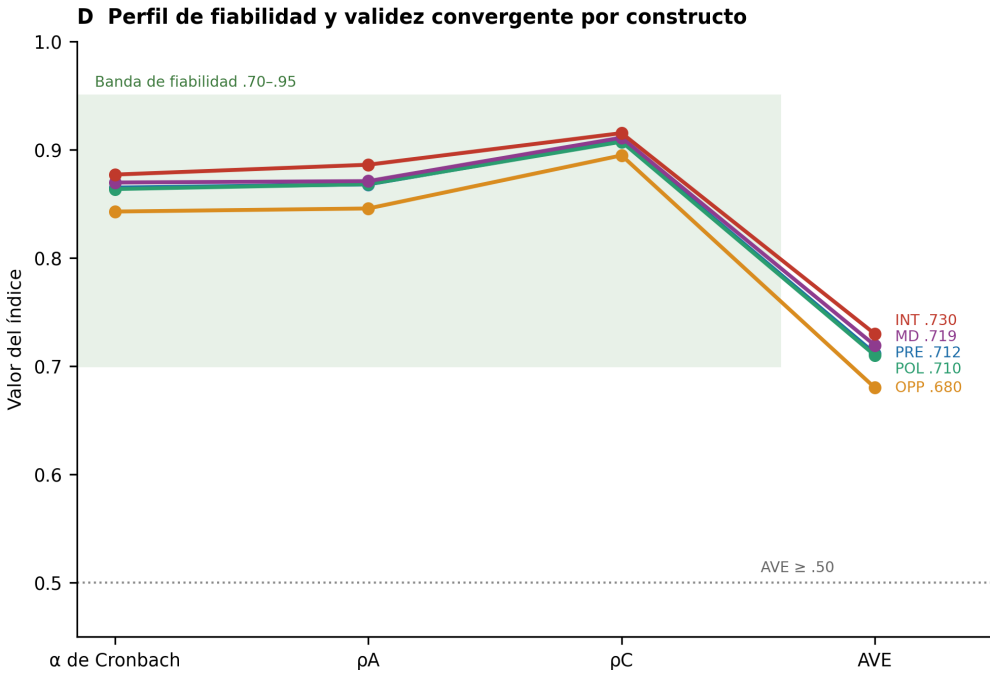


Nota. N = 310. Panel B: distribución de cada carga externa en 5000 remuestreos bootstrap (densidad sombreada), con la estimación puntual (círculo vacío) y el intervalo percentil del 95 % (línea vertical); las cargas oscilaron entre .784 y .884 y el límite inferior más bajo de los intervalos fue .718, por encima de la línea de referencia de .708.

C Razones heterorrasgo-monorrasgo entre constructos



Nota. N = 310. Panel C: razón heterorrasgo-monorrasgo (HTMT) de cada par de constructos (grosor y color de la arista), con el límite superior de su intervalo bootstrap del 95 % entre corchetes y la raíz cuadrada de la AVE de cada constructo; la razón más alta fue .577 (MD–INT; límite superior .667), por debajo de .85.



Nota. N = 310. Panel D: α de Cronbach, ρA, ρC y AVE de cada constructo frente a la banda de fiabilidad .70–.95 y el umbral de AVE de .50; la AVE osciló entre .680 y .730.

Tabla 6-6

Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis

Hipótesis	Camino	β	EE	t	p	IC 95 % (BCa)	f^2	VIF interno	Decisión
H1a	PRE → MD	.267	.049	5.50	< .001	[.168, .356]	.091	1.04	Apoyada
H1b	PRE → INT	.169	.049	3.44	< .001	[.072, .265]	.040	1.15	Apoyada
H2a	OPP → MD	.147	.059	2.47	.013	[.040, .271]	.026	1.09	Apoyada
H2b	OPP → INT	.198	.052	3.80	< .001	[.089, .293]	.049	1.27	Apoyada
H3a	POL → MD	−.294	.056	−5.26	< .001	[−.392, −.172]	.103	1.11	Apoyada
H3b	POL → INT	−.072	.048	−1.52	.130	[−.164, .025]	.007	1.23	No apoyada
H4	MD → INT	.352	.050	7.01	< .001	[.250, .447]	.149	1.32	Apoyada
Control	USE → INT	.125	.051	2.46	.014	[.025, .222]	.021	1.20	—

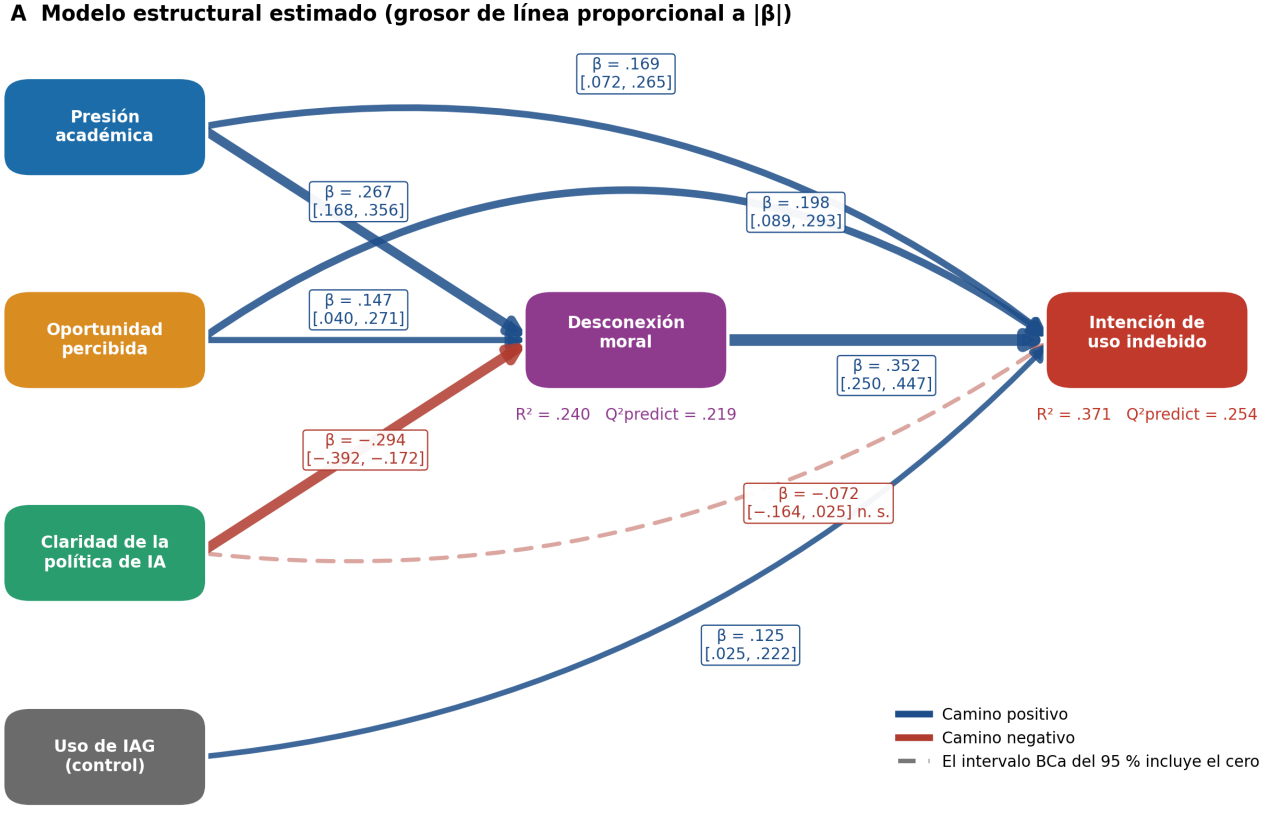
Nota: β = coeficiente de camino estandarizado; EE = error estándar bootstrap (5000 remuestreos); los valores p son bilaterales; BCa = intervalo corregido por sesgo y acelerado. Las hipótesis direccionales se apoyan cuando el intervalo del 95 % excluye el cero en la dirección hipotetizada. R^2(MD) = .240; R^2(INT) = .371.

6.4.4 Poder explicativo y predictivo

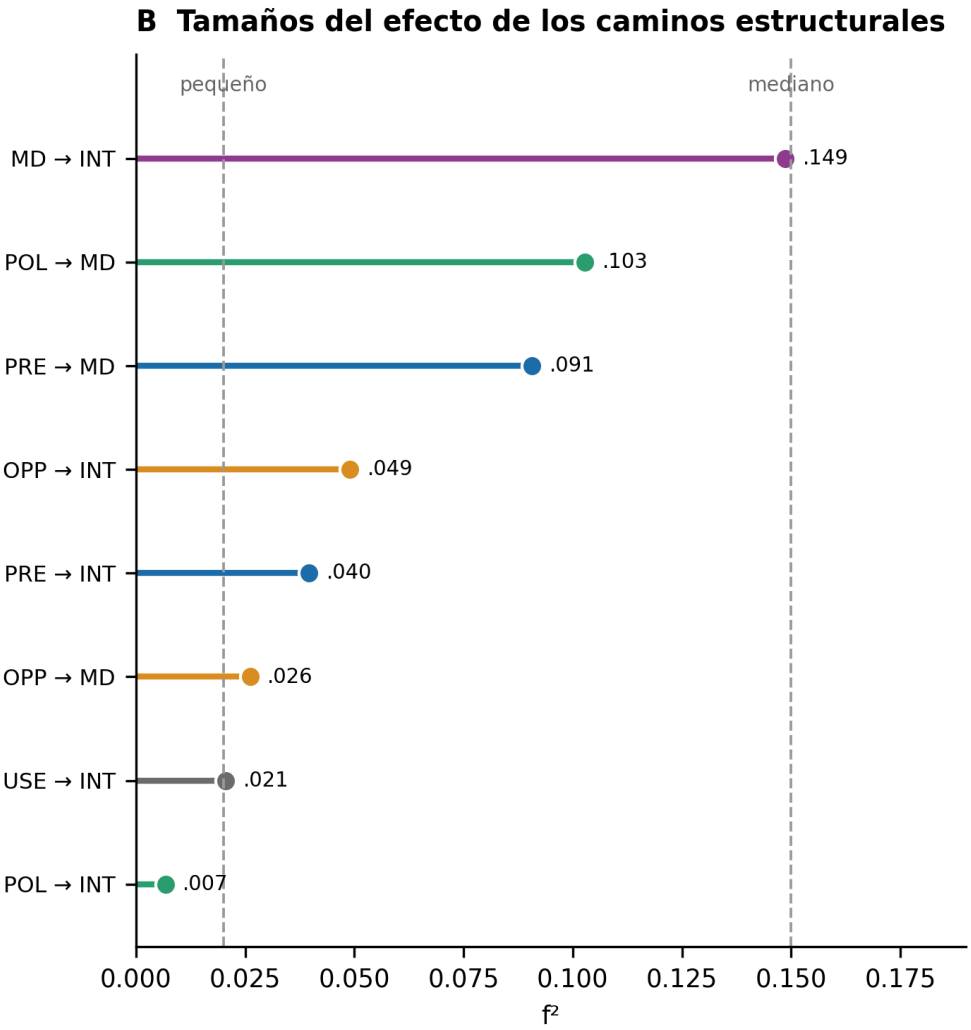
El modelo explicó el 24.0 % de la varianza de la desconexión moral y el 37.1 % de la de la intención (Tabla 6-7). Fuera de muestra, el Q^2 predict fue positivo en los ocho indicadores endógenos y en ambos constructos, y PLS-SEM produjo menor RMSE y menor MAE que el modelo lineal de referencia en los ocho indicadores, lo que indica un poder predictivo alto. El SRMR fue .049 para el modelo saturado y .050 para el estimado, ambos inferiores a .08.

Figura 6-2

Modelo estructural, tamaños del efecto y distribuciones bootstrap conjuntas de los efectos indirectos específicos sobre la intención de uso indebido

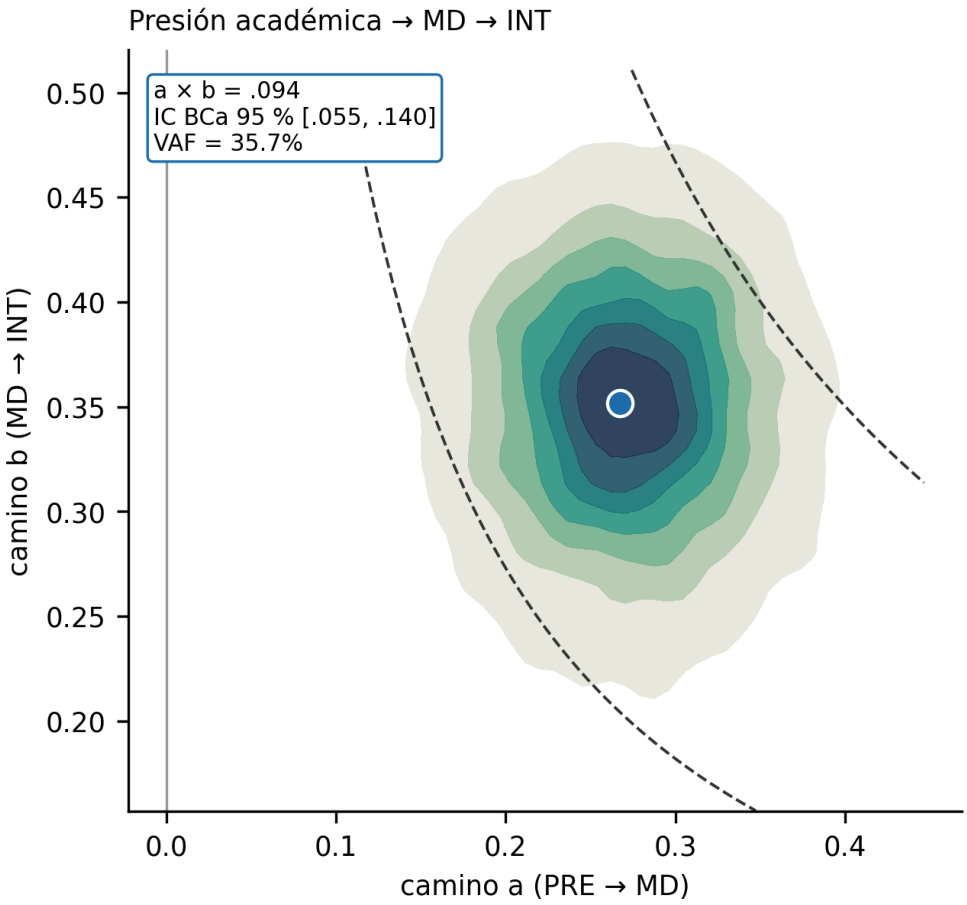


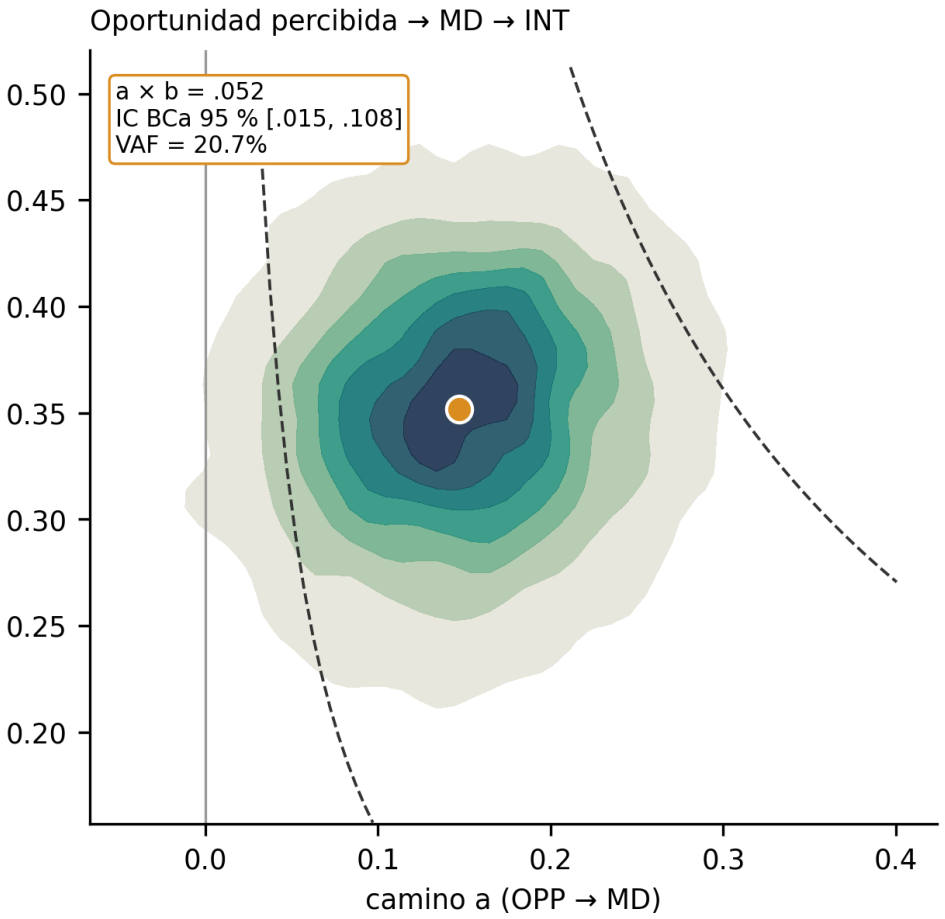
Nota. N = 310; inferencia con 5000 remuestreos bootstrap e intervalos del 95 % corregidos por sesgo y acelerados (BCa). Panel A: modelo PLS-SEM estimado; el grosor de cada línea es proporcional a |β|, las líneas azules son caminos positivos y las rojas negativos, y la línea discontinua señala el único camino cuyo intervalo incluye el cero (POL → INT, β = −.072, IC 95 % [−.164, .025]); MD → INT β = .352, IC 95 % [.250, .447]; PRE → MD β = .267, IC 95 % [.168, .356]; OPP → MD β = .147, IC 95 % [.040, .271]; POL → MD β = −.294, IC 95 % [−.392, −.172]; PRE → INT β = .169, IC 95 % [.072, .265]; OPP → INT β = .198, IC 95 % [.089, .293]; USE → INT β = .125, IC 95 % [.025, .222]. Bajo cada constructo endógeno se muestran R^2 y Q^2predict (MD: R^2 = .240, Q^2predict = .219; INT: R^2 = .371, Q^2predict = .254).

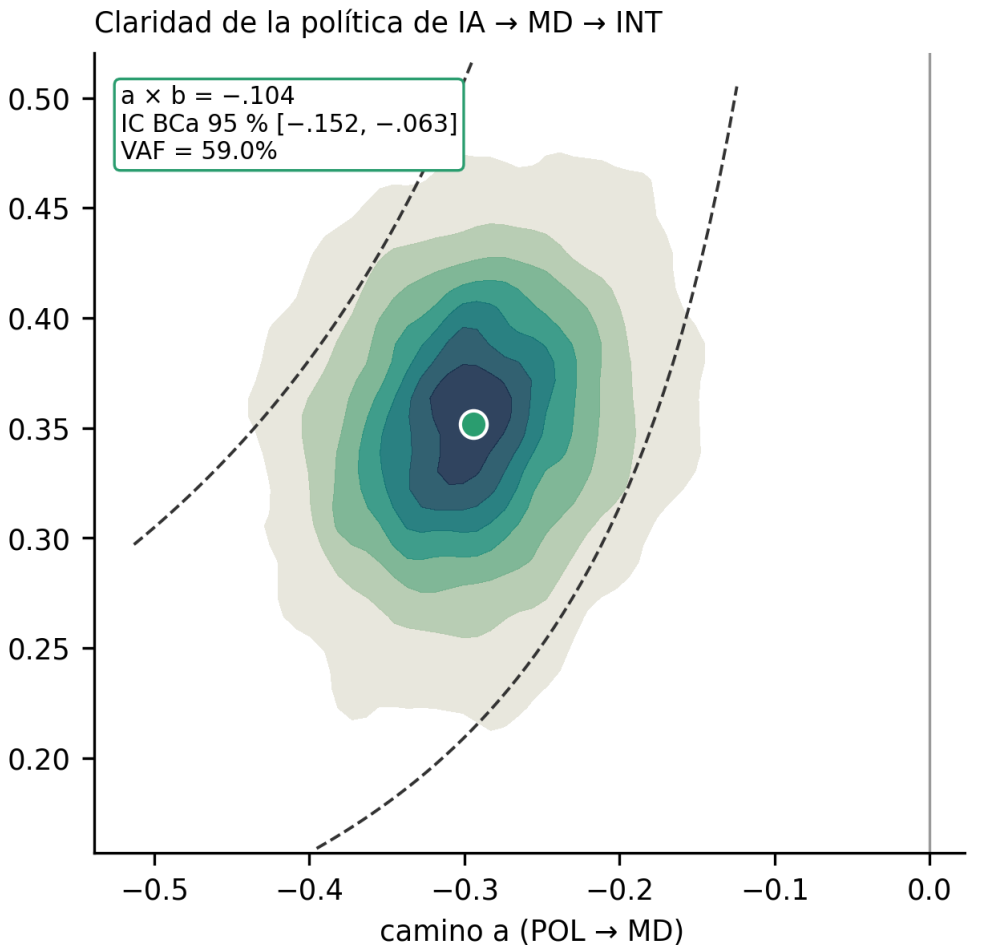


Nota. N = 310; inferencia con 5000 remuestreos bootstrap e intervalos del 95 % corregidos por sesgo y acelerados (BCa). Panel B: f^2 de cada camino estructural con los umbrales convencionales de efecto pequeño (.02) y mediano (.15); los valores oscilaron entre .007 (POL → INT) y .149 (MD → INT).

C Distribución bootstrap conjunta de los caminos a y b de cada efecto indirecto específico




Oportunidad percibida → MD → INT
a × b = .052
IC BCa 95 % [.015, .108]
VAF = 20.7%
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.0
0.1
0.2
0.3
0.4
camino a (OPP → MD)



Nota. N = 310; inferencia con 5000 remuestreos bootstrap e intervalos del 95 % corregidos por sesgo y acelerados (BCa). Panel C: contornos de densidad de los 5000 pares bootstrap del camino a (antecedente → MD) y del camino b (MD → INT) de cada efecto indirecto específico; el círculo relleno es la estimación puntual, las curvas discontinuas son las combinaciones cuyo producto a × b iguala los límites inferior y superior del intervalo BCa del 95 %, y el recuadro informa el efecto indirecto y la varianza explicada por la mediación (VAF): PRE → MD → INT = .094, IC 95 % [.055, .140], VAF = 35.7 %; OPP → MD → INT = .052, IC 95 % [.015, .108], VAF = 20.7 %; POL → MD → INT = −.104, IC 95 % [−.152, −.063], VAF = 59.0 %.

Tabla 6-7

Poder explicativo y capacidad predictiva fuera de muestra (PLSpredict)

Panel	Objetivo	R^2		Q^2 predict	RMSE		MAE	
		[IC 95 %]	ajustado		(PLS)	(LM)	(PLS)	(LM)
A. Constructos	Desconexión moral (MD)	.240 [.164, .343]	.232	.219	—	—	—	—
	Intención de uso indebido (INT)	.371 [.295, .469]	.361	.254	—	—	—	—
B. Indicadores	MD1	—	—	.147	1.194	1.215	0.974	0.993
	MD2	—	—	.172	1.174	1.194	0.942	0.962
	MD3	—	—	.149	1.161	1.195	0.923	0.951
	MD4	—	—	.162	1.154	1.182	0.936	0.956
	INT1	—	—	.219	1.108	1.131	0.873	0.898
	INT2	—	—	.175	1.157	1.176	0.950	0.955
	INT3	—	—	.183	1.137	1.157	0.916	0.927
	INT4	—	—	.160	1.159	1.192	0.920	0.940

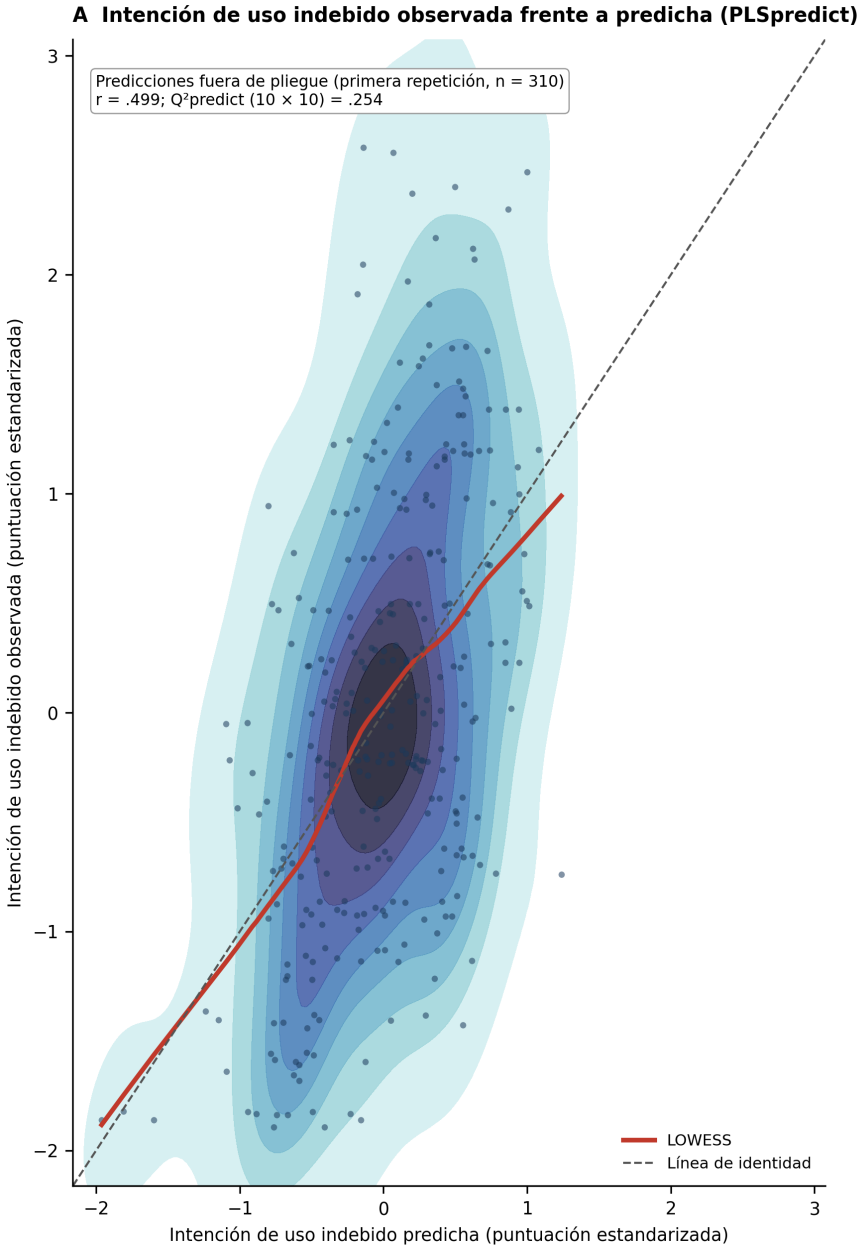
Nota: Q^2predict, RMSE y MAE proceden de una validación cruzada de 10 pliegues repetida 10 veces. LM = modelo lineal de referencia que regresa cada indicador endógeno sobre los indicadores de sus constructos exógenos antecedentes. Un RMSE y un MAE menores indican mejor predicción.

6.4.5 Mediación

Los tres efectos indirectos específicos a través de la desconexión moral fueron significativos, con apoyo a H5a–c (Tabla 6-8; Figura 6-2, panel C). En la presión y la oportunidad, los efectos directo e indirecto fueron significativos y del mismo signo (mediación parcial complementaria). En la claridad de la política, el efecto indirecto fue significativo y el directo no, lo que corresponde a mediación solo indirecta, y su efecto total sobre la intención fue de –.176.

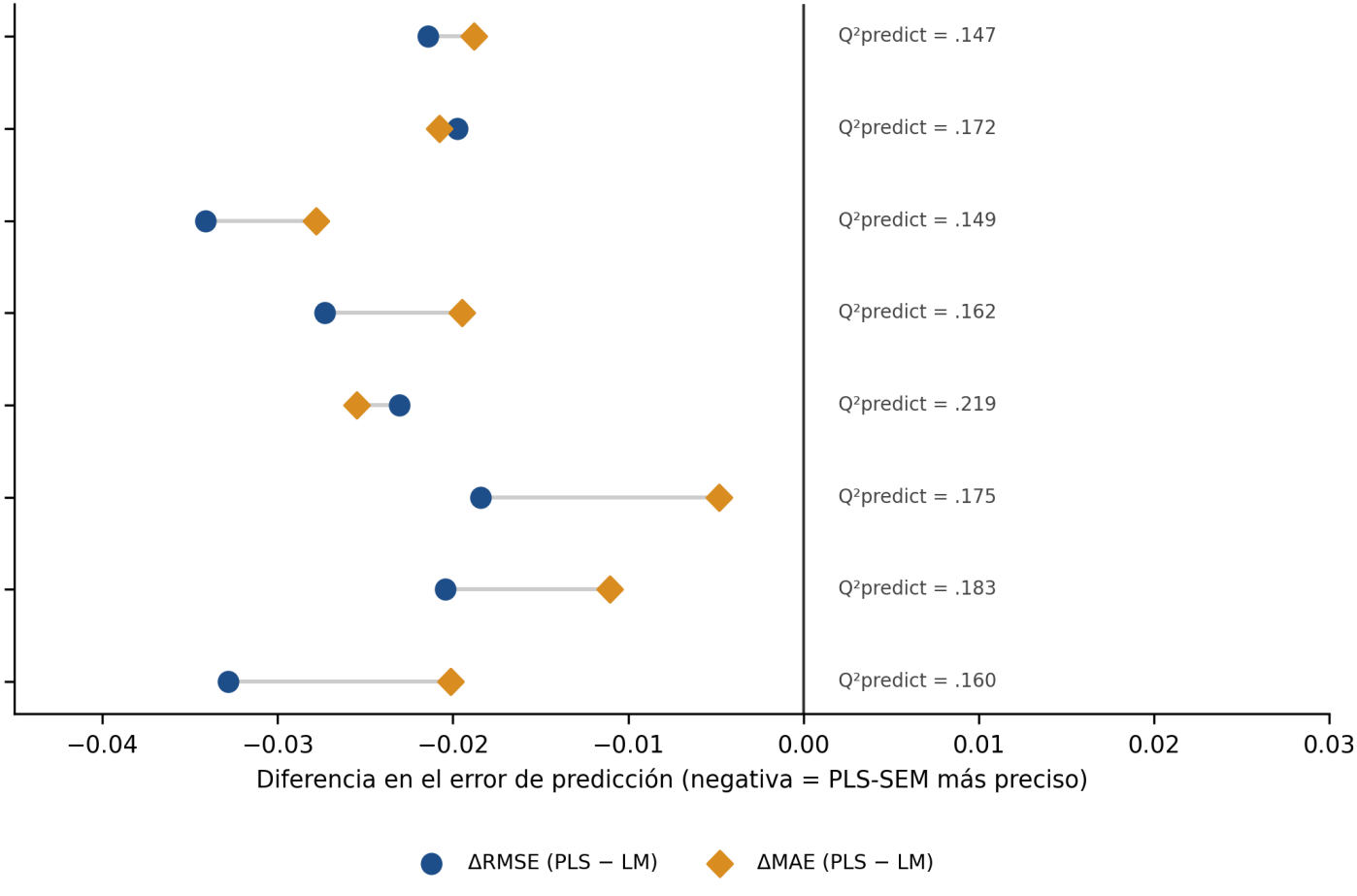
Figura 6-3

Capacidad predictiva fuera de muestra y varianza explicada del modelo

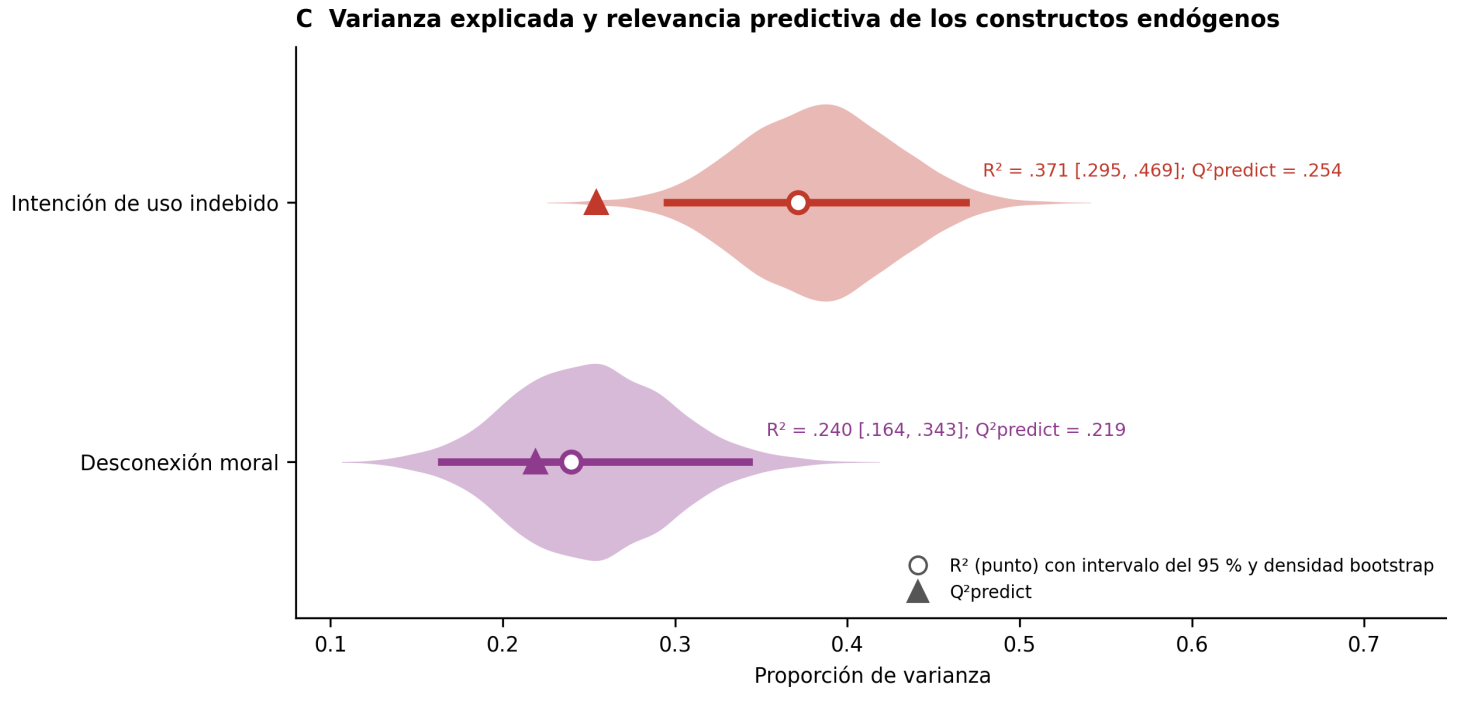


Nota. N = 310; PLSpredict con 10 pliegues y 10 repeticiones. Panel A: puntuaciones estandarizadas observadas frente a predichas de la intención de uso indebido en las predicciones fuera de pliegue de la primera repetición (n = 310), con contornos de densidad, una curva LOWESS y la línea de identidad; r = .499, y el Q²predict del procedimiento 10 × 10 fue .254.

B Errores de predicción fuera de muestra de PLS-SEM frente al modelo lineal de referencia



Nota. N = 310; PLSpredict con 10 pliegues y 10 repeticiones. Panel B: diferencia en la raíz del error cuadrático medio (RMSE) y en el error absoluto medio (MAE) entre PLS-SEM y el modelo lineal de referencia para cada indicador de la desconexión moral y de la intención de uso indebido, con el Q^2predict a la derecha; las 8 diferencias en RMSE (−.034 a −.018) y en MAE (−.028 a −.005) son negativas, es decir, PLS-SEM predijo cada indicador con más precisión que el modelo de referencia. Panel C: R^2 de la desconexión moral (.240, IC 95 % [.164, .343]) y de la intención de uso indebido (.371, IC 95 % [.295, .469]) con sus densidades bootstrap, y Q^2predict de cada constructo (triángulos).



Nota. N = 310; PLSpredict con 10 pliegues y 10 repeticiones. Panel C: R^2 de la desconexión moral (.240, IC 95 % [.164, .343]) y de la intención de uso indebido (.371, IC 95 % [.295, .469]) con sus densidades bootstrap, y Q^2predict de cada constructo (triángulos).

Tabla 6-8

Descomposición de los efectos de los antecedentes sobre la intención de uso indebido en directos, indirectos y totales

Hipótesis	Antecedente	Efecto sobre INT	Estimado	EE	t	p	IC 95 % (BCa)	VAF	Tipo de mediación
H5a	Presión académica (PRE)	Directo	.169	.049	3.44	< .001	[.072, .265]	—	Parcial complementaria
		Indirecto vía MD	.094	.022	4.31	< .001	[.055, .140]	35.7 %	
		Total	.263	.048	5.50	< .001	[.166, .352]	—	
H5b	Oportunidad percibida (OPP)	Directo	.198	.052	3.80	< .001	[.089, .293]	—	Parcial complementaria
		Indirecto vía MD	.052	.023	2.24	.025	[.015, .108]	20.7 %	
		Total	.249	.058	4.30	< .001	[.131, .360]	—	
H5c	Claridad de la política de IA (POL)	Directo	−.072	.048	−1.52	.130	[−.164, .025]	—	Solo indirecta (total)
		Indirecto vía MD	−.104	.023	−4.57	< .001	[−.152, −.063]	59.0 %	
		Total	−.176	.051	−3.47	< .001	[−.272, −.073]	—	

Nota: Los efectos indirectos son productos de caminos estandarizados (antecedente → MD × MD → INT). EE e intervalos BCa de 5000 remuestreos bootstrap con aceleración por jackknife. VAF = varianza explicada (efecto indirecto/efecto total). Tipos de mediación: parcial complementaria (efectos directo e indirecto significativos y del mismo signo) y solo indirecta (efecto indirecto significativo y efecto directo no significativo).

6.4.6 Robustez

Ningún término cuadrático fue significativo (Tabla 6-9), por lo que la especificación lineal resultó adecuada. Las puntuaciones de POL ($D = .119$, $p < .001$) y MD ($D = .079$, $p < .001$) se apartaron de la normalidad, lo que identifica sus términos de cópula; ninguno fue significativo.

Tabla 6-9

Comprobaciones de robustez: efectos no lineales y endogeneidad

Comprobación	Término	Ecuación	Estimado	EE	t	p	IC 95 % (BCa)
No linealidad (término cuadrático en dos etapas)	MD²	INT	.035	.044	0.81	.418	[–.051, .118]
	PRE²	MD	–.005	.045	–0.11	.911	[–.092, .083]
	OPP²	MD	.061	.060	1.01	.313	[–.059, .174]
Endogeneidad (cópula gaussiana)	Cópula (POL)	INT	.056	.184	0.31	.760	[–.304, .412]
	Cópula (POL)	MD	.097	.195	0.50	.620	[–.324, .442]
	Cópula (MD)	INT	–.237	.476	–0.50	.618	[–.969, .928]

Nota: Los efectos indirectos son productos de caminos estandarizados (antecedente → MD × MD → INT). EE e intervalos BCa de 5000 remuestreos bootstrap con aceleración por jackknife. VAF = varianza explicada (efecto indirecto/efecto total). Tipos de mediación: parcial complementaria (efectos directo e indirecto significativos y del mismo signo) y solo indirecta (efecto indirecto significativo y efecto directo no significativo).

6.4.7 Invarianza de medida y análisis multigrupo

La invarianza composicional se cumplió en todos los constructos: ninguna correlación entre compuestos quedó por debajo del cuantil del 5 % de su distribución de permutación (valor p mínimo = .201; Tabla 6-10). Las medias y log-varianzas de los compuestos no difirieron entre grupos (valor p mínimo = .261), por lo que se estableció invarianza plena y los datos pudieron analizarse agrupados. Respecto de la PI1, ninguna de las ocho diferencias de camino ni de las tres diferencias de efectos indirectos fue significativa, antes (p mínimo = .062) o después del ajuste de Holm (p ajustado mínimo = .684). El R^2 de la intención fue .321 en el grupo STEM y .437 en el no STEM.

Tabla 6-10

Invarianza de medida de los compuestos (MICOM) y análisis multigrupo por grupo disciplinar

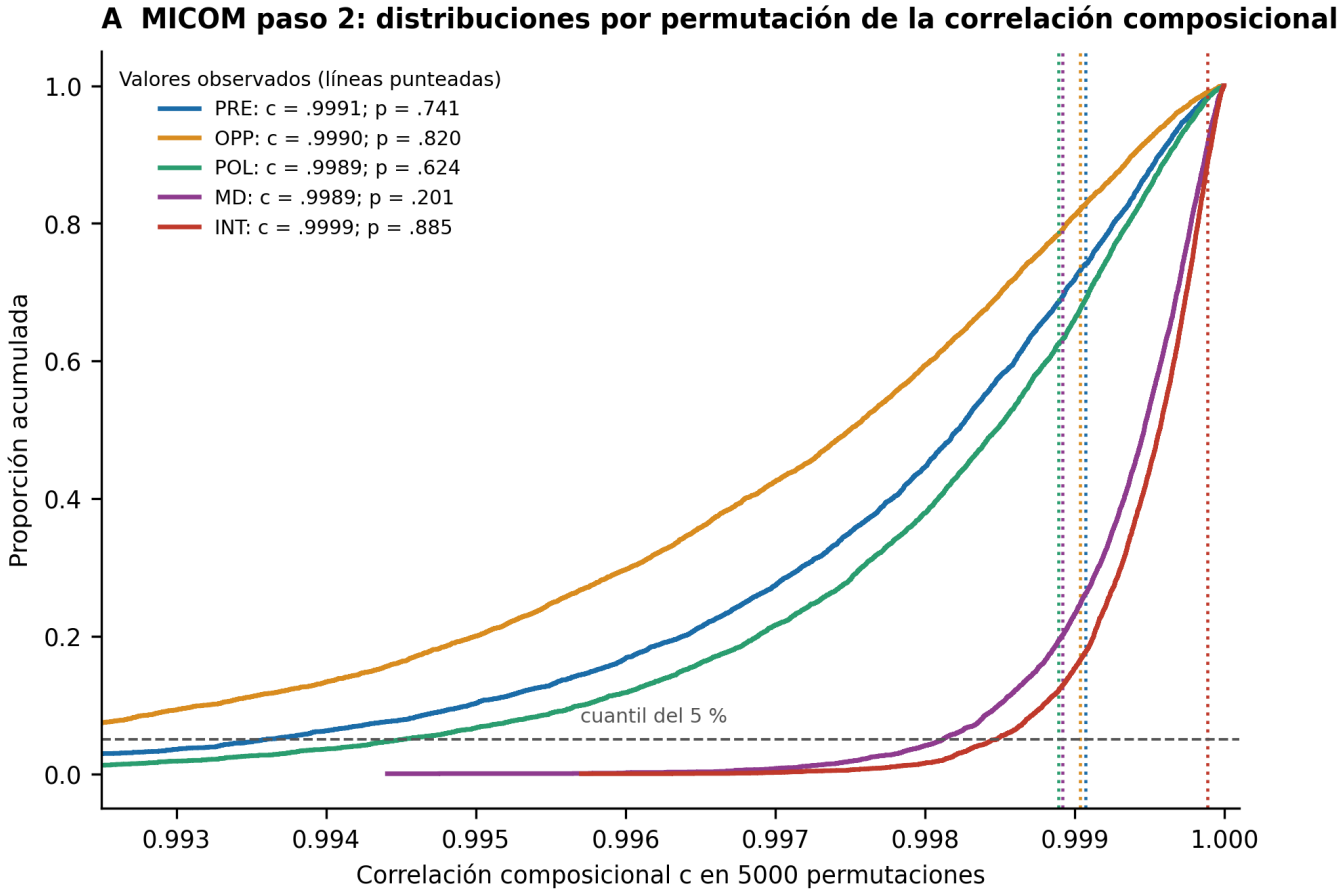
	Panel	Constructo o camino	Estadístico	STEM	No STEM	Diferencia	p	
							de permuta	ajustado (Holm)
A.	MICOM paso 2	PRE	c = .9991 (cuantil 5 % .9936)	—	—	—	.741	—
		OPP	c = .9990 (cuantil 5 % .9914)	—	—	—	.820	—
		POL	c = .9989 (cuantil 5 % .9945)	—	—	—	.624	—
		MD	c = .9989 (cuantil 5 % .9981)	—	—	—	.201	—
		INT	c = .9999 (cuantil 5 % .9985)	—	—	—	.885	—
	MICOM paso 3	PRE	Δmedia = .103 [−.216, .228]; Δlog var = −.051 [−.292, .304]	—	—	—	.365; .751	—
		OPP	Δmedia = .036 [−.219, .218]; Δlog var = −.036 [−.285, .281]	—	—	—	.747; .795	—
		POL	Δmedia = .037 [−.217, .223]; Δlog var = −.229 [−.399, .402]	—	—	—	.748; .261	—
		MD	Δmedia = −.002 [−.214, .217]; Δlog var = −.085 [−.294, .289]	—	—	—	.984; .563	—
		INT	Δmedia = .077 [−.220, .222]; Δlog var = −.037 [−.287, .287]	—	—	—	.490; .798	—

Panel	Constructo o camino	Estadístico	STEM	No STEM	Diferencia	p	
						de permuta	ajustado (Holm)
B. AMG	PRE → MD	β	.366	.182	.183	.062	.684
	OPP → MD	β	.126	.168	−.042	.734	1.000
	POL → MD	β	−.248	−.333	.085	.457	1.000
	PRE → INT	β	.102	.241	−.139	.166	1.000
	OPP → INT	β	.217	.174	.043	.685	1.000
	POL → INT	β	−.082	−.057	−.026	.793	1.000
	MD → INT	β	.288	.422	−.134	.199	1.000
	USE → INT	β	.146	.105	.041	.692	1.000
	PRE → MD → INT	a × b	.105	.077	.028	.517	1.000
	OPP → MD → INT	a × b	.036	.071	−.035	.480	1.000
	POL → MD → INT	a × b	−.072	−.141	.069	.125	1.000

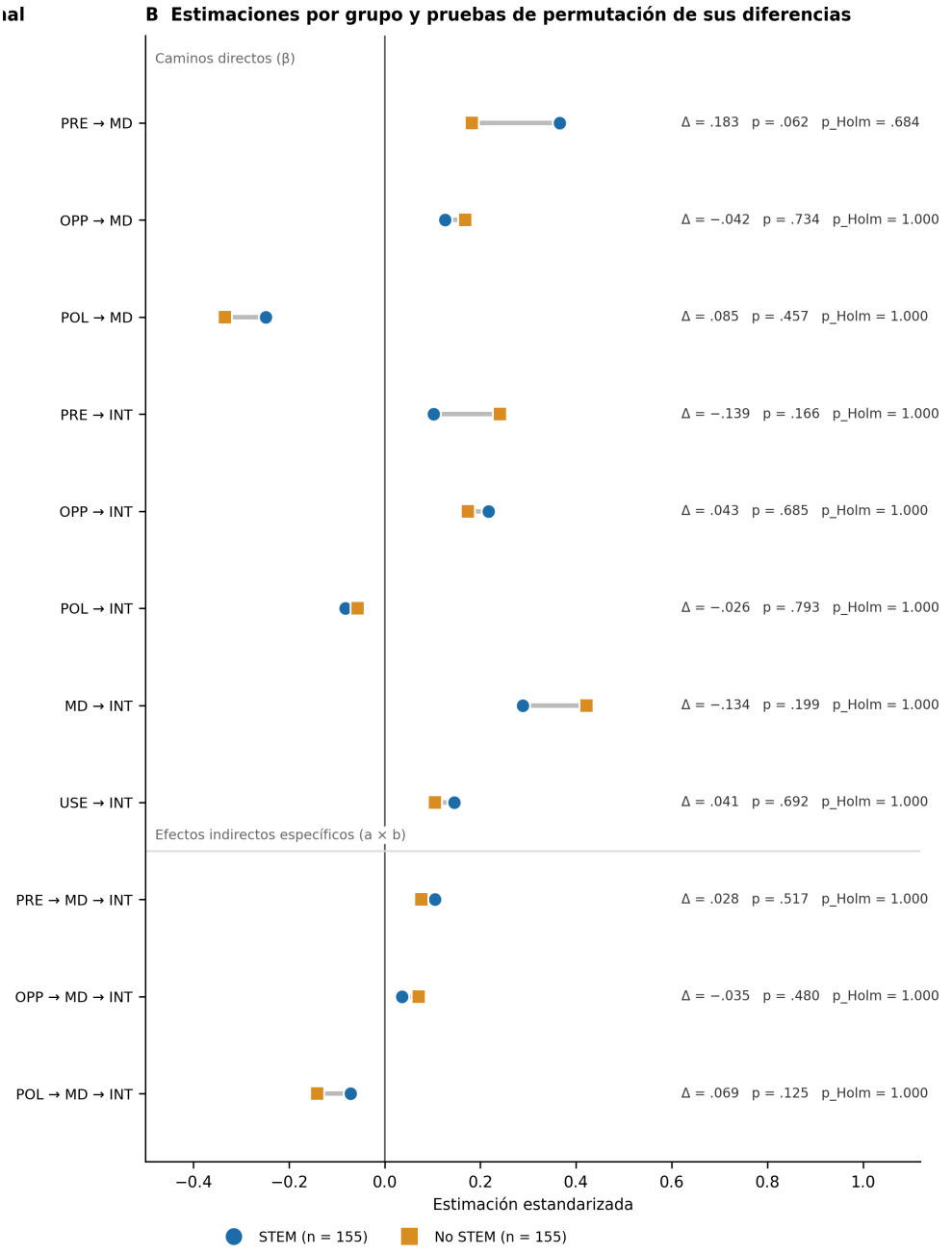
Nota: STEM n = 155; no STEM n = 155. El paso 2 contrasta la invarianza composicional (c = correlación entre las puntuaciones compuestas calculadas con los pesos de cada grupo; hay invarianza cuando c no es inferior al cuantil del 5 % de su distribución de permutación). El paso 3 contrasta la igualdad de medias y log-varianzas de los compuestos (intervalos de permutación del 95 % entre corchetes; valores p de medias; varianzas). El panel B presenta coeficientes de camino y efectos indirectos por grupo; valores p de 5000 permutaciones, ajustados por Holm en los 11 contrastes. AMG = análisis multigrupo.

Figura 6-4

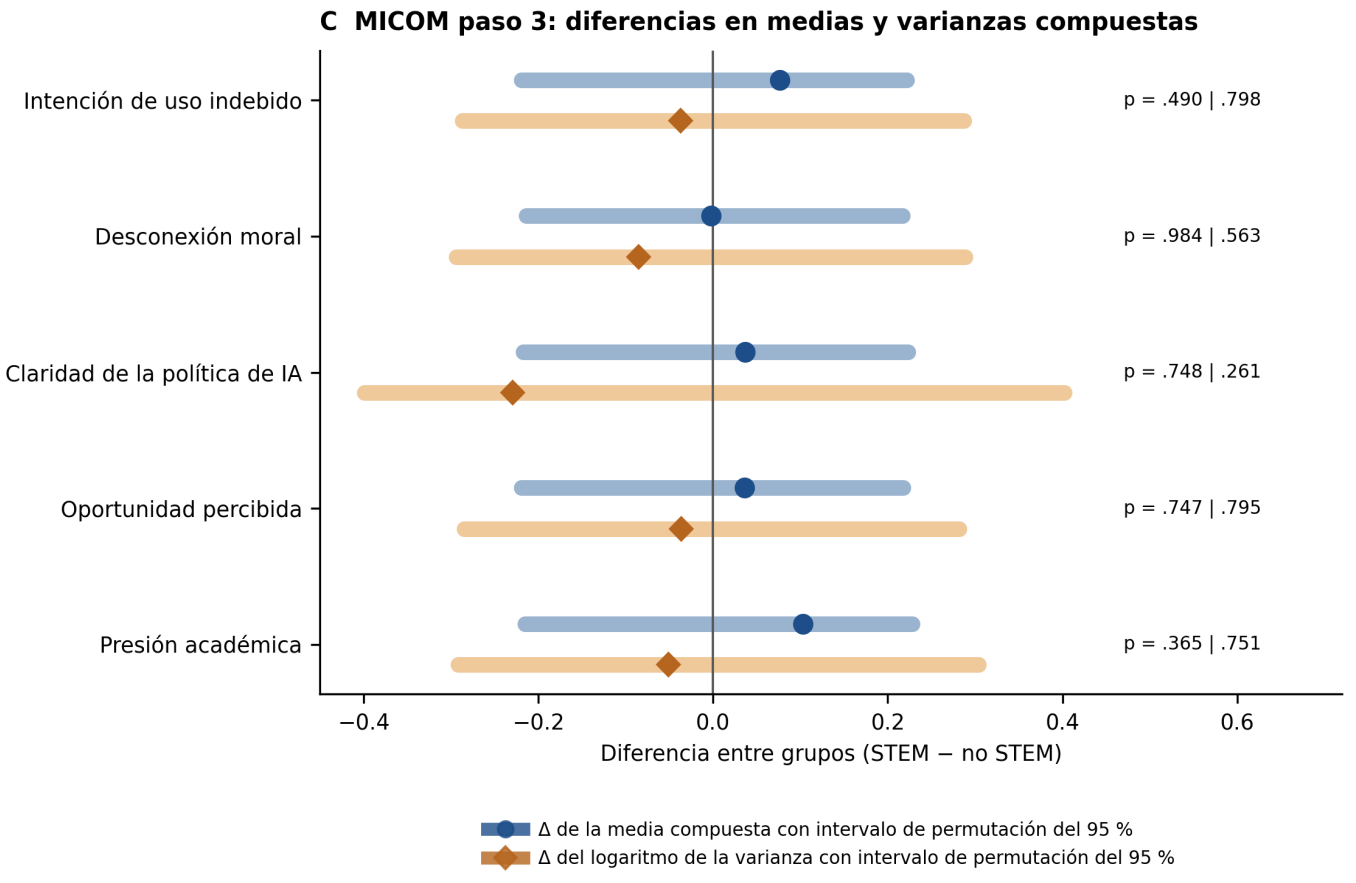
Invarianza de medida de los compuestos y análisis multigrupo entre estudiantes de carreras STEM y no STEM.



Nota. STEM n = 155; no STEM n = 155; 5000 permutaciones. Panel A: distribución acumulada empírica de la correlación composicional c bajo permutación para cada constructo, con el valor observado (línea punteada) y el nivel del 5 % (línea discontinua); el eje horizontal empieza en .9925. Valores observados: PRE c = .9991 (cuantil del 5 % .9936, p = .741); OPP c = .9990 (cuantil del 5 % .9914, p = .820); POL c = .9989 (cuantil del 5 % .9945, p = .624); MD c = .9989 (cuantil del 5 % .9981, p = .201); INT c = .9999 (cuantil del 5 % .9985, p = .885).



Nota. Panel B: coeficientes de camino estandarizados y efectos indirectos específicos por grupo (círculos, STEM; cuadrados, no STEM) con la diferencia, el valor p de permutación y el valor p ajustado por Holm en los 11 contrastes; ninguna diferencia fue significativa (p mínimo = .062, PRE → MD; p ajustado por Holm mínimo = .684).



Nota. STEM n = 155; no STEM n = 155; 5000 permutaciones. Panel C: diferencias entre grupos en las medias compuestas y en el logaritmo de las varianzas, con sus intervalos de permutación del 95 % y sus valores p (medias | varianzas); todas las diferencias quedan dentro de sus intervalos (p mínimo = .261).

6.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre desconexión moral e intención, sino que la claridad de la política de IA, el antecedente con la relación negativa más intensa con la desconexión moral, careciera de efecto directo sobre la intención y, aun así, mostrara un efecto total negativo y significativo transmitido casi por completo por ese mecanismo. Un análisis que contrastara únicamente el camino directo habría concluido que las normas son irrelevantes; la descomposición sugiere lo contrario: las normas claras parecen operar sobre las razones que el estudiante puede esgrimir, no sobre la conducta en abstracto. Esta lectura es consistente con la ausencia de efecto de las políticas institucionales documentada en el modelo más reciente del triángulo del fraude (Rahman et al., 2026) y ofrece una explicación plausible de esa inconsistencia, aunque el diseño transversal impide afirmar que ese sea el único mecanismo.

La interpretación gana verosimilitud al contrastarla con la literatura institucional. Los análisis de políticas describen textos que tratan la IA como una amenaza tecnológica más, sin redefinir la autoría aceptable (Perkins y Roe, 2024), y un estudio de misiones universitarias identifica la brecha entre los ideales comunicados y los efectivamente implementados (Taylor y LaCroix, 2026). El imperativo ético que se destila de las primeras respuestas institucionales, que los trabajos reflejen el conocimiento individual adquirido y que las personas conserven la responsabilidad moral de las faltas asistidas por IA (Dabis y Csáki, 2024), es precisamente lo que niega el ítem de desplazamiento de la responsabilidad empleado aquí («si la universidad no lo prohíbe explícitamente, no puede culparse al estudiante»). Que la claridad de las reglas reduzca esa justificación encaja con la propuesta de que los principios éticos y pedagógicos deben ser transparentes y operar en el nivel de la tarea (Nguyen, 2025) y con los modelos de gobernanza que, desde un contexto mexicano, buscan articular la innovación educativa con la integridad (Mejía y Góngora, 2025). Desde una perspectiva crítica, los modelos que entran la racionalización como

tercer predictor paralelo, y en los que todos los factores del triángulo resultan significativos (Alshurafat et al., 2024), no pueden distinguir entre un efecto propio de la oportunidad y uno mediado por las justificaciones que esta habilita; la estructura secuencial aquí contrastada permite hacerlo y muestra que, en presión y oportunidad, ambas vías coexisten. En consecuencia, la mediación complementaria sugiere que la presión y la oportunidad no solo suministran razones, sino que ejercen un empuje residual sobre la intención que la desconexión moral no agota, una proporción que fue menor en la oportunidad que en la presión.

Por otra parte, la lectura de la presión se apoya en la evidencia experimental según la cual las situaciones de alta presión incrementan la probabilidad y la magnitud de la trampa, sin que la interacción con la orientación de metas sea significativa (Yin et al., 2025). Los resultados actuales sugieren un canal adicional: la carga excesiva y las consecuencias del fracaso no solo empujan, sino que proporcionan el argumento con el que la falta se vuelve defendible, tal como recoge el primer ítem de desconexión moral. La oportunidad, en cambio, mostró un efecto más débil sobre la desconexión moral que sobre la intención directa, lo que es compatible con que la facilidad de uso y la baja probabilidad de sanción operen más como condición habilitante que como justificación; no obstante, la comparación entre estos dos caminos es descriptiva y no se contrastó formalmente. La estabilidad de las tasas de trampa observada en secundaria tras la aparición de los agentes conversacionales (Chen et al., 2026) invita a pensar que la disponibilidad tecnológica por sí sola no explica la conducta, algo coherente con que el peso del modelo recaiga en la desconexión moral y no en el acceso, si bien la diferencia de nivel educativo y de diseño exige cautela al trasladar esa evidencia.

La centralidad de la desconexión moral concuerda con la evidencia previa. Su papel como vía de transmisión de la conducta indebida en el trabajo (Ogunfowora et al., 2022), la fuerza de su asociación con la deshonestidad académica (Sirca y Billen, 2024; Flutti et al., 2026) y su función en cadena entre rasgos de personalidad y uso no ético de la IAG, donde el maquiavelismo y la psicopatía se relacionaron con el uso problemático

solo de manera indirecta, sin efectos directos ni totales significativos (Kırcaburun y Özdemir, 2026), dibujan un patrón en el que la desconexión actúa como eslabón y no como mero correlato. Algo semejante puede afirmarse de las emociones morales: la culpa anticipada, más frecuente que el remordimiento posterior en una muestra de personal académico (Vassallo, 2026), sugiere que las autosanciones se activan antes de la acción y que las justificaciones medidas aquí podrían ser el modo de silenciarlas, aunque la población y el objeto de ese estudio difieren de los nuestros y el vínculo es una conjetura.

En cuanto a la generalidad, la ausencia de diferencias entre estudiantes STEM y no STEM, tras establecer la invarianza de medida, sugiere que el mecanismo es una propiedad de cómo el estudiantado razona sobre el uso indebido y no de las culturas de evaluación de cada disciplina. Esto complementa los resultados multigrupo en los que las diferencias surgieron en actitudes y percepciones antes que en los vínculos entre ellas (Ofem et al., 2025). Conviene matizar que el R^2 de la intención fue numéricamente mayor en el grupo no STEM y que la diferencia de PRE → MD rozó la significación antes del ajuste, de modo que la potencia por grupo, calibrada para caminos de magnitud moderada, no permite descartar diferencias pequeñas. El contexto peruano, además, aporta una prueba en un sistema donde la regulación específica sobre integridad ante la IAG escasea (Cabrera et al., 2026); la percepción estudiantil de que el uso es, en parte, una práctica de apropiación parcial, como la que declara una proporción elevada de estudiantes de contabilidad en Brasil, que copia parcialmente las respuestas o las reescribe (Alves et al., 2025), es coherente con una zona gris normativa en la que las justificaciones prosperan.

6.5.1 Implicaciones teóricas y prácticas

Desde el punto de vista teórico, reformular la racionalización como mediador convierte el triángulo del fraude en un modelo de proceso y lo conecta con la teoría de la agencia moral (Bandura et al., 1996), lo que ayuda a entender por qué una misma presión objetiva conduce a unos estudiantes, y no a otros, al uso indebido. Para las universidades, los

resultados indican que las estrategias centradas en la detección abordan un solo antecedente, además poco fiable (Cotton et al., 2024; Eke, 2023), y que el punto de apalancamiento es la desconexión moral. Las directrices institucionales, a menudo generales y heterogéneas (Jin et al., 2025; An et al., 2025), deberían traducirse en instrucciones por evaluación que especifiquen los usos permitidos, la forma de declararlos y las consecuencias del abuso. Dado que la presión actuó por vía directa e indirecta, el diseño de evaluaciones con menos solapamientos de plazos y menos entregas únicas de alto impacto es también una medida de integridad. La educación en integridad puede, por último, abordar explícitamente las justificaciones medidas, y el desarrollo del carácter y el liderazgo ético propuestos para el uso responsable de ChatGPT (Crawford et al., 2023) ofrece un marco para ello.

6.5.2 Limitaciones y líneas futuras

El diseño transversal no establece precedencia temporal, y diseños longitudinales permitirían contrastar la dirección de la relación entre desconexión moral e intención. La intención es un antecedente proximal de la conducta, no la conducta, y futuros trabajos podrían vincularla con indicadores observados preservando el anonimato. El muestreo por cuotas no probabilístico restringe la inferencia a los estudiantes analizados, y los ítems de intención, pese a su formulación hipotética, pueden estar afectados por deseabilidad social. Finalmente, el 24.0 % de varianza explicada de la desconexión moral deja margen para incorporar diferencias individuales, como los rasgos antagónicos (Kırcaburun y Özdemir, 2026), o factores contextuales adicionales.

6.6 CONCLUSIONES

La intención de usar indebidamente la IAG está impulsada por la desconexión moral, que transmite los efectos de la presión académica, la oportunidad percibida y la claridad de la política institucional de IA. La presión y la oportunidad incrementan la intención de forma directa y a través de las justificaciones que hacen disponibles, mientras que la

claridad de la política la reduce al disminuir esas justificaciones, sin un componente directo significativo. Este patrón resuelve una inconsistencia de la literatura: el efecto nulo de las políticas sobre la conducta, observado cuando solo se contrasta el camino directo, es compatible con una influencia sustantiva ejercida por la vía cognitiva.

El modelo de medida fue fiable y válido, el modelo estructural predijo la intención fuera de muestra mejor que una referencia lineal y las relaciones se mantuvieron entre estudiantes de áreas STEM y no STEM, lo que respalda la generalidad del mecanismo en el contexto peruano, escasamente representado en la evidencia previa. La aportación del estudio es teórica, al recolocar la racionalización como mediador del triángulo del fraude, y metodológica, al sumar evidencia predictiva y de invarianza.

Las implicaciones son directas. Las instituciones protegen la integridad académica en la era de la IAG menos mediante la detección del uso indebido que mediante la eliminación de las justificaciones cognitivas que lo hacen aceptable, comenzando por reglas claras, formuladas a nivel de cada evaluación, sobre los usos permitidos, su declaración y sus consecuencias, y por diseños de evaluación que alivien la presión. Los límites del estudio, su carácter transversal, el muestreo por cuotas y la medición de intenciones y no de conductas, invitan a replicarlo con diseños longitudinales y con indicadores observados en otros sistemas universitarios.

6.7 REFERENCIAS BIBLIOGRÁFICAS

Abbas, M., Khan, T., & Jam, F. (2025). Avoid excessive usage: Examining the motivations and outcomes of generative artificial intelligence usage among students. *Journal of Academic Ethics*, *23*(4), 2423–2442. https://doi.org/10.1007/s10805-025-09659-3

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Al-Nuaimi, M., & Uzun, A. (2023). Socio-cognitive determinants of plagiarism intentions among university students during emergency online learning: Integrating emotional, motivational, and moral factors into theory of planned behavior. *Cogent Psychology*, *10*(1), Article 2168362. https://doi.org/10.1080/23311908.2023.2168362

Alshurafat, H., Al Shbail, M., Hamdan, A., Al-Dmour, A., & Ensour, W. (2024). Factors affecting accounting students' misuse of ChatGPT: An application of the fraud triangle theory. *Journal of Financial Reporting and Accounting*, *22*(2), 274–288. https://doi.org/10.1108/jfra-04-2023-0182

Alves, M., Silva, C., & Bonfim, M. (2025). ChatGPT e desonestidade acadêmica: percepção dos estudantes de contabilidade sobre o seu uso. *Revista Gestão Universitária na América Latina - GUAL*, 133–152. https://doi.org/10.5007/1983-4535.2024.e99115

An, Y., Yu, J., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, *22*(1), Article 10. https://doi.org/10.1186/s41239-025-00507-3

Bandura, A., Barbaranelli, C., Caprara, G., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364–374. https://doi.org/10.1037/0022-3514.71.2.364

Becker, J., Proksch, D., & Ringle, C. (2022). Revisiting Gaussian copulas to handle endogenous regressors. *Journal of the Academy of Marketing Science*, *50*(1), 46–66. https://doi.org/10.1007/s11747-021-00805-y

Cabrera, J., Llanos, R., & Bonilla, L. (2026). Generative AI and the assurance of academic integrity in the context of South American higher education. *Frontiers in Education, 11*, 1796556. https://doi.org/10.3389/feduc.2026.1796556

Chen, R., Lee, V., Kuo, A., Pope, D., & Miles, S. (2026). Cheating in the second year of generative AI chatbots: A follow-up study on high school student cheating behaviors. *Educational Technology Research and Development, 74*(2), 639–667. https://doi.org/10.1007/s11423-026-10587-1

Cotton, D., Cotton, P., & Shipway, J. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice, 20*(3). https://doi.org/10.53761/1.20.3.02

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*(1), Article 1006. https://doi.org/10.1057/s41599-024-03526-z

Duong, N., Huynh, Q., Nguyen, H., Tran, T., & Thai, N. (2026). Academic dishonesty in the digital age: Toward sustainable educational integrity through TPB. *TEM Journal, 15*(3), 2576. https://doi.org/10.18421/tem153-47

Eke, D. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, *13*, Article 100060. https://doi.org/10.1016/j.jrt.2023.100060

Fajt, B., & Schiller, E. (2025). ChatGPT in academia: University students' attitudes towards the use of ChatGPT and plagiarism. *Journal of Academic Ethics*, *23*(3), 1363–1382. https://doi.org/10.1007/s10805-025-09603-5

Flutti, E., Travaini, G., Roma, P., & Ongis, M. (2026). Neutralization techniques and academic dishonesty among Italian university students. *Frontiers in Psychology*, *17*, Article 1766612. https://doi.org/10.3389/fpsyg.2026.1766612

Gruenhagen, J., Sinclair, P., Carroll, J., Baker, P., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, *7*, Article 100273. https://doi.org/10.1016/j.caeai.2024.100273

Hair, J., Hult, G., Ringle, C., Sarstedt, M., Danks, N., & Ray, S. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook*. Springer. https://doi.org/10.1007/978-3-030-80519-7

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405–431. https://doi.org/10.1108/IMR-09-2014-0304

Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence, 8,* Article 100348. https://doi.org/10.1016/j.caeai.2024.100348

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261. https://doi.org/10.1111/isj.12131

Kofinas, A., Tsay, C., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology, 56*(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Kırcaburun, K., & Özdemir, P. (2026). Moral disengagement and unethical generative AI use as the chain mediators between antagonistic personality and problematic generative AI use. *Behavioral Sciences, 16*(4), Article 500. https://doi.org/10.3390/bs16040500

Lingán, S., Dominguez, S., Cárdenas, S., & Baca, Y. (2026). Construction of a scale to measure moral disengagement associated with plagiarism in young university students. *International Journal of Adolescence and Youth, 31*(1), Article 2629235. https://doi.org/10.1080/02673843.2026.2629235

Mejía, R., & Góngora, E. (2025). Inteligencia artificial generativa e integridad académica en educación superior: hacia un modelo de gobernanza institucional para la innovación educativa. *Reencuentro. Análisis de problemas universitarios, 37*(89), 249–298. https://doi.org/10.24275/uamxoc-dcsh/reencuentro/202589-12

Nguyen, K. (2025). The use of generative AI tools in higher education: Ethical and pedagogical principles. *Journal of Academic Ethics, 23*(3), 1435–1455. https://doi.org/10.1007/s10805-025-09607-1

Ofem, U., Owan, V., Iyam, M., Udeh, M., Anake, P., & Ovat, S. (2025). Students' perceptions, attitudes and utilisation of ChatGPT for academic dishonesty: Multigroup analyses via PLS-SEM. *Education and Information Technologies, 30*(1), 159–187. https://doi.org/10.1007/s10639-024-12850-5

Ogunfowora, B., Nguyen, V., Steel, P., & Hwang, C. (2022). A meta-analytic investigation of the antecedents, theoretical correlates, and consequences of moral disengagement at work. *Journal of Applied Psychology, 107*(5), 746–775. https://doi.org/10.1037/apl0000912

Ortiz, S., & Blahopoulou, J. (2025). Chat or cheat? Academic dishonesty, risk perceptions, and ChatGPT usage in higher education students. *Social Psychology of Education, 28*(1), Article 113. https://doi.org/10.1007/s11218-025-10080-2

Perkins, M., & Roe, J. (2024). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy, 37*(3), 633–653. https://doi.org/10.1057/s41307-023-00323-2

Pérez, A., de Andrés, J., Arias, M., & Padilla, G. (2026). Ethical judgments and weakness of will in the misuse of generative AI: The case of academic essay production among higher education students. *Innovative Higher Education, 51*(2), 453–479. https://doi.org/10.1007/s10755-026-09885-6

Rahman, M., Nur, F., Liza, B., Hasan, M., & Rahman, M. (2026). Exploring the moderating effect of academic support on Bangladeshi accounting students' use of ChatGPT for academic dishonesty using fraud triangle theory. *Journal of Applied Research in Higher Education*. Advance online publication. https://doi.org/10.1108/jarhe-11-2025-0939

Sarstedt, M., Hair, J., Pick, M., Liengaard, B., Radomir, L., & Ringle, C. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing, 39*(5), 1035–1064. https://doi.org/10.1002/mar.21640

Schoemann, A., Boulton, A., & Short, S. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Schuberth, F., Rademaker, M., & Henseler, J. (2023). Assessing the overall fit of composite models estimated by partial least squares path

modeling. *European Journal of Marketing, 57*(6), 1678–1702. https://doi.org/10.1108/EJM-08-2020-0586

Sirca, C., & Billen, E. (2024). Predicting academic dishonesty: The role of psychopathic traits, perception of academic dishonesty, moral disengagement and motivation. *Journal of Academic Ethics, 22*(3), 489–503. https://doi.org/10.1007/s10805-024-09506-x

Song, N. (2024). Higher education crisis: Academic misconduct with generative AI. *Journal of Contingencies and Crisis Management, 32*(1), Article e12532. https://doi.org/10.1111/1468-5973.12532

Strzelecki, A. (2024). Students' acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology. *Innovative Higher Education, 49*(2), 223–245. https://doi.org/10.1007/s10755-023-09686-1

Taylor, T., & LaCroix, T. (2026). Purpose before policy: Academic integrity, generative AI, and rhetorical stance. *Higher Education.* Advance online publication. https://doi.org/10.1007/s10734-026-01706-1

Torres, J., Rivera, D., & Beltrán, A. (2026). Generative AI Adoption and Self-Reported Academic Integrity Risk: A Student Taxonomy in Latin American Higher Education. *Education Sciences, 16*(9), 1397. https://doi.org/10.3390/educsci16091397

Vassallo, D. (2026). The AI guilt complex: Moral emotions and ethical dilemmas in academic technology adoption. *Journal of Academic Ethics, 24*(1), Article 53. https://doi.org/10.1007/s10805-026-09726-3

Yeung, L., Chow, D., Wong, P., & Lau, S. (2026). In ChatGPT they trust: A study of students' perceptions and misuse of ChatGPT in higher education. *AI and Ethics, 6*(1), Article 11. https://doi.org/10.1007/s43681-025-00855-w

Yilmaz, A. (2026). The psychological mechanisms and behavioral determinants of academic integrity in the age of artificial intelligence. *Frontiers in Psychology, 17*, Article 1853790. https://doi.org/10.3389/fpsyg.2026.1853790

Yin, X., Lai, Q., & Guo, Y. (2025). Drivers of academic dishonesty: Situational pressures versus student motivation. *International Journal of Adolescence and Youth, 30*(1), Article 2560649. https://doi.org/10.1080/02673843.2025.2560649

Yusuf, A., Pervin, N., & Román, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education, 21*(1), Article 21. https://doi.org/10.1186/s41239-024-00453-6

CAPÍTULO VII

EXPOSICIÓN A PATRONES TEXTILES ANDINOS, VISUALIZACIÓN ESPACIAL Y PENSAMIENTO GEOMÉTRICO EN ESTUDIANTES DE SECUNDARIA DEL PERÚ: UN MODELO DE MEDIACIÓN CON INVARIANZA DE MEDICIÓN

Judith Esther Gavidia Medrano

Universidad Nacional Hermilio Valdizán

Arnulfo Ortega Mallqui

Universidad Nacional Hermilio Valdizán

Resumen

La etnomatemática sostiene que los artefactos culturales portan conocimiento geométrico, pero la evidencia disponible compara programas y no identifica el mecanismo cognitivo que conecta el contacto con esos artefactos con el pensamiento geométrico escolar. Este estudio contrastó, con los textiles andinos como caso de prueba, un modelo en el que la visualización espacial media la asociación entre la exposición a patrones textiles y el pensamiento geométrico. Se empleó un diseño cuantitativo, transversal y explicativo con 407 estudiantes de secundaria de instituciones públicas del Perú, seleccionados por conveniencia, cuyo tamaño se fijó mediante potencia Monte Carlo. Los tres constructos se midieron con escalas de autoinforme de tres ítems y se estimó un modelo de ecuaciones estructurales basado en covarianzas, con bootstrap de 5 000 réplicas para el efecto indirecto e invarianza de medición por lengua materna. El modelo se ajustó muy bien (CFI = .999; RMSEA = .013; SRMR = .0185) y la invarianza métrica y escalar se sostuvo. La exposición se asoció con el pensamiento geométrico de forma directa (β = .493) e indirecta (a·b = .193; IC 95 % [.031, .352]), con una mediación parcial que transmitió el 28.1 % del efecto total. La validez discriminante entre exposición y visualización espacial requiere cautela, pues el criterio de Fornell-Larcker no se cumplió. Los resultados sugieren que los artefactos con estructura geométrica operan como entornos de práctica de transformaciones mentales, con implicaciones transferibles a otros patrimonios artesanales.

Palabras clave: etnomatemática; textiles andinos; visualización espacial; pensamiento geométrico; ecuaciones estructurales; invarianza de medición.

Abstract

Ethnomathematics holds that cultural artifacts carry geometric knowledge, yet the available evidence compares programs and does not identify the cognitive mechanism linking contact with such artifacts to school geometric thinking.

Using Andean textiles as a test case, this study examined a model in which spatial visualization mediates the association between exposure to textile patterns and geometric thinking. A quantitative, cross-sectional, explanatory design was applied to 407 secondary students from public schools in Peru, selected by convenience, with sample size set through Monte Carlo power analysis. The three constructs were measured with three-item self-report scales, and a covariance-based structural equation model was estimated, with a 5,000-replication bootstrap for the indirect effect and measurement invariance by native language. The model fit very well (CFI = .999; RMSEA = .013; SRMR = .0185), and metric and scalar invariance held. Exposure was associated with geometric thinking both directly (β = .493) and indirectly (a·b = .193; 95% CI [.031, .352]), a partial mediation transmitting 28.1% of the total effect. Discriminant validity between exposure and spatial visualization warrants caution, as the Fornell-Larcker criterion was not met. The results suggest that artifacts with geometric structure act as practice environments for mental transformations, with implications transferable to other craft heritages.

Keywords: *ethnomathematics; Andean textiles; spatial visualization; geometric thinking; structural equation modeling; measurement invariance.*

7.1 INTRODUCCIÓN

La enseñanza de la geometría escolar se apoya, en la mayoría de los sistemas educativos, en representaciones descontextualizadas de figuras y transformaciones, mientras que muchas comunidades producen artefactos cuya elaboración exige razonar con simetrías, traslaciones, rotaciones y teselaciones. La etnomatemática sostiene que esas prácticas culturales portan conocimiento matemático propio y que su reconocimiento en el aula puede sostener el aprendizaje formal (D'Ambrosio, 1985). El programa ha adquirido además una dimensión política, pues valorar y respetar los saberes y haceres matemáticos locales constituye uno de los objetivos de los procesos decolonizadores en la educación matemática (Rosa y Orey, 2025). La evidencia acumulada respalda la promesa pedagógica de este enfoque: las síntesis metaanalíticas de intervenciones etnomatemáticas

reportan efectos de d = 0.94 en 11 estudios (Wirawan et al., 2024) y de d = 1.15 en 15 estudios (Hartoyo et al., 2025) sobre el rendimiento matemático. Sin embargo, esa evidencia procede de comparaciones entre programas y no identifica el mecanismo cognitivo por el cual el contacto con artefactos culturales con estructura geométrica se traduce en pensamiento geométrico.

El textil es uno de los artefactos andinos en los que esa estructura se manifiesta con mayor densidad. En el Perú, el tejido se comprende como una práctica integral transmitida entre generaciones, según lo documenta el estudio cualitativo con tejedoras de la comunidad de Acchahuata, en Cusco (Palomino et al., 2024), y su repertorio gráfico entronca con los tocapus incaicos, cuya herencia perdura en el pallay de los tejidos de Chinchero (Valle, 2022). La continuidad de estos repertorios no es exclusiva del país: se registra en los símbolos puruhá de las alfombras de Guano, en Ecuador (Maldonado et al., 2026), y en la indumentaria del Cacique Tisaleo (Flores y Otáñez, 2025). Los análisis etnomatemáticos han examinado la elaboración de un textil andino peruano desde el pensamiento complejo (Bonilla, 2023), los saberes del tejido en lana de la comunidad yanacona de Caquiona, en Colombia, frente a su desarticulación con la matemática escolar (Morales, 2024), los conceptos presentes en la confección del maki poncho puruhá (Atupaña et al., 2025) y las conexiones del arte textil mazahua y hñähñu con las actividades matemáticas de Bishop (Gilsdorf, 2024). Un estudio transcultural de tecnologías de fibra identifica, a su vez, acciones que corresponden a conceptos numéricos y geométricos fundamentales (Mendoza et al., 2026), mientras que la reflexión sobre el pensamiento geométrico andino ha señalado el significado particular de la diagonal para los pueblos originarios y ha cuestionado interpretaciones geométricas impuestas desde fuera (Park, 2022).

La investigación sobre cognición espacial ofrece un candidato explícito para el mecanismo ausente. Las habilidades espaciales son maleables: su entrenamiento mejora tanto la visualización espacial (g = .49) como el rendimiento matemático (g = .28) en 29 estudios con diseño pre-post controlado (Hawes et al., 2022), y la síntesis de esa evidencia ha llevado

a postular un papel causal del pensamiento espacial en las matemáticas, todavía ausente de los currículos (Gilligan et al., 2022). En el Perú, un cuasiexperimento con estudiantes de secundaria de Lima sugiere que la educación artística incide en el pensamiento abstracto-matemático (Davila, 2026), lo cual apunta a que la práctica con formas y artefactos visuales puede transferirse al razonamiento matemático. En la teoría de los niveles de van Hiele (1986), el pensamiento geométrico progresa desde el reconocimiento visual de las formas hacia el análisis de sus propiedades, de modo que una práctica que ejercite la manipulación mental de configuraciones podría facilitar ese tránsito.

Si la exposición cotidiana a artefactos con transformaciones geométricas ejercita la manipulación mental de formas, la visualización espacial debería transmitir parte de su asociación con el pensamiento geométrico. Esta hipótesis de mediación no ha sido contrastada con un modelo de medición latente que separe la varianza del constructo del error de medida, ni se ha verificado que los instrumentos funcionen de manera equivalente en estudiantes con lengua materna originaria y con castellano, condición indispensable para comparar aulas culturalmente diversas.

El presente estudio aborda ese vacío con los textiles andinos como caso de prueba. Su contribución es de alcance general: propone y contrasta un modelo del mecanismo cognitivo que conecta el conocimiento geométrico codificado en un artefacto cultural con el pensamiento geométrico escolar, transferible a cualquier sistema educativo con patrimonio artesanal (tejido, cestería, cerámica o arquitectura vernácula), y aporta evidencia de invarianza de medición entre grupos lingüísticos. El Perú, donde la tradición textil andina mantiene vigencia en la vida cotidiana de los estudiantes de secundaria, ofrece el contexto empírico para esa contrastación.

La pregunta de investigación es: ¿en qué medida la exposición a patrones textiles andinos se asocia con el pensamiento geométrico de estudiantes de secundaria del Perú, y qué parte de esa asociación se transmite a través de la visualización espacial? Se contrastan cuatro hipótesis.

1) **H1**: la exposición a patrones textiles andinos se asocia positiva y directamente con el pensamiento geométrico.

2) **H2**: la exposición se asocia positivamente con la visualización espacial.

3) **H3**: la visualización espacial se asocia positivamente con el pensamiento geométrico.

4) **H4**: la exposición ejerce un efecto indirecto positivo sobre el pensamiento geométrico a través de la visualización espacial.

7.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

1) El marco principal es la etnomatemática, que concibe las matemáticas como un conjunto de modos de explicar, medir, clasificar e inferir que cada grupo cultural desarrolla en sus prácticas (D'Ambrosio, 1985). Desde esta perspectiva, el textil andino no es una ilustración del contenido escolar, sino un sistema de conocimiento en el que la composición de motivos mediante reflexiones, giros y repeticiones constituye una forma de razonamiento geométrico. La exposición a estos patrones se define aquí como la frecuencia de contacto, la participación en su elaboración y el reconocimiento de su estructura simétrica. Esta definición reúne tres facetas del vínculo con el artefacto, la observación, la práctica y el reconocimiento analítico, y evita reducir la exposición a una mera presencia pasiva de objetos en el entorno.

2) El segundo marco es la teoría de los niveles de pensamiento geométrico de van Hiele (1986), que describe una progresión desde el reconocimiento visual de figuras hacia el análisis de sus propiedades, las relaciones entre ellas y la deducción. El reconocimiento de simetrías y transformaciones ocupa los niveles iniciales de esa progresión, de modo que la familiaridad con artefactos que las exhiben de forma sistemática puede facilitar el paso del reconocimiento perceptivo al análisis de propiedades. De ahí se deriva H1: la exposición a patrones textiles debería asociarse directamente con el pensamiento geométrico, con independencia de cualquier vía intermedia.

3) El tercer marco, la cognición espacial, aporta el mecanismo. La visualización espacial es la capacidad de representar y transformar mentalmente objetos y configuraciones, y se fortalece con la práctica (Hawes et al., 2022). La decodificación de un textil exige justamente rotar, reflejar y descomponer motivos, por lo que la exposición debería asociarse con una mayor visualización espacial (H2). A su vez, las transformaciones mentales son el sustrato operativo de las tareas geométricas de simetría, congruencia y teselación (Hawes et al., 2022), lo que fundamenta H3; la evidencia que atribuye al pensamiento espacial un papel causal en el aprendizaje matemático refuerza esa expectativa (Gilligan et al., 2022). La conjunción de H2 y H3 implica que parte de la asociación entre exposición y pensamiento geométrico se transmita a través de la visualización espacial (H4), sin excluir la vía directa de H1; el modelo anticipa, por tanto, una mediación parcial y no una mediación total.

7.3 MATERIALES Y MÉTODOS

7.3.1 Diseño

Se empleó un diseño cuantitativo, no experimental, transversal y explicativo, con la exposición a patrones textiles andinos como constructo exógeno, la visualización espacial como mediador y el pensamiento geométrico como constructo endógeno final. Los datos corresponden al año escolar 2025. La unidad de análisis fue el estudiante y el modelo se estimó en un solo nivel; la agrupación en aulas e instituciones no se modeló.

7.3.2 Participantes y tamaño muestral

Participaron estudiantes de educación secundaria de instituciones públicas del Perú, reclutados por conveniencia en las instituciones que aceptaron participar; la selección no fue probabilística y los resultados describen al grupo analizado. El tamaño muestral se determinó a priori mediante tres cálculos de potencia ($\alpha = .05$; $1 - \beta = .80$), uno por inferencia, y se adoptó

el mayor. Para el ajuste global, el modelo con nueve indicadores dispone de 45 momentos y estima 21 parámetros libres (gl = 24); la prueba de ajuste cercano (H0: $\varepsilon \leq .05$ frente a H1: $\varepsilon = .08$) requirió N = 375 (potencia = .801; MacCallum et al., 1996). Para la ruta b, con b = .14 ($f^2 = .0200$) y la prueba F no central, se requirieron N = 395 (potencia = .800). Para el efecto indirecto a·b se aplicó potencia Monte Carlo con intervalo de confianza Monte Carlo del 95 % (Fossum y Montoya, 2023), con 1 000 réplicas por tamaño candidato, 2 000 extracciones por réplica y a = .24, b = .14 y $c' = .10$, lo que alcanzó una potencia de .817 con N = 407. Los valores de a y b son el extremo conservador del rango de efectos publicado, r = .14–.50 (obtenido con $r = d/\sqrt{(d^2 + 4)}$), que procede de 55 estudios: el entrenamiento espacial mejoró la visualización espacial (g = .49; r = .24) y el rendimiento matemático (g = .28; r = .14) en 29 estudios (Hawes et al., 2022), y los enfoques etnomatemáticos produjeron efectos de d = 0.94 a 1.15 (r = .43–.50) en 26 estudios (Wirawan et al., 2024; Hartoyo et al., 2025). La muestra quedó fijada en N = 407, el mínimo que garantiza una potencia de al menos .80 en el análisis más exigente.

7.3.3 Instrumentos

Los tres constructos se midieron con escalas de autoinforme de tres ítems, con respuesta Likert de cinco puntos, construidas para este estudio y validadas mediante análisis factorial confirmatorio en la propia muestra (Tabla 7-1). La Escala de Exposición a Patrones Textiles Andinos (EPTA) recogió la frecuencia de contacto, la participación en la elaboración y el reconocimiento de motivos simétricos. La Escala de Visualización Espacial Autopercibida (EVEA) recogió la rotación mental de figuras tridimensionales, el plegado mental de papel y la identificación de patrones espaciales ocultos, en la línea de los trabajos que articulan visualización espacial y razonamiento geométrico (Merma et al., 2024). La Escala de Pensamiento Geométrico Autopercibido (EPGA) recogió la identificación de propiedades de figuras y cuerpos, la resolución de problemas de simetría, traslación y rotación, y la argumentación en teselación, alineadas con la competencia geométrica del Currículo Nacional de la Educación Básica del

Perú. Se registraron cuatro covariables, utilizadas solo en las regresiones de robustez: la memoria de trabajo visoespacial (tarea de bloques de Corsi), la lengua materna (1 = originaria; 0 = castellano), la escolaridad discontinua (1 = con interrupción) y el sexo (registro dicotómico de matrícula). La lengua materna definió además los grupos del análisis de invarianza.

Tabla 7-1

Operacionalización de los constructos del modelo

Constructo	Rol en el modelo	Instrumento	Ítems	Escala
Exposición a patrones textiles andinos	Exógeno (X)	EPTA (autoinforme, construida para este estudio)	3	Likert de 5 puntos
Pensamiento geométrico	Endógeno final (Y)	EPGA (autoinforme; ítems alineados con la competencia geométrica del Currículo Nacional)		
Visualización espacial	Mediador (M)	EVEA (autoinforme; ítems inspirados en rotación mental y plegado de papel)		

Nota: Los tres constructos son reflectivos y se miden con tres ítems cada uno. Las propiedades psicométricas (λ, α, ω, AVE) y la validez discriminante se reportan en las Tablas 7-2 y 7-3. Los tres constructos son reflectivos y se miden con tres ítems cada uno. Las propiedades psicométricas (λ, α, ω, AVE) y la validez discriminante se reportan en las Tablas 7-2 y 7-3.

7.3.4 Análisis de datos

El modelo se estimó como un modelo de ecuaciones estructurales basado en covarianzas por máxima verosimilitud con la librería semopy. El modelo de medición se evaluó con las cargas estandarizadas, la fiabilidad (α de Cronbach y ω), la varianza extraída promedio (AVE) y la validez discriminante mediante la ratio heterotrait-monotrait (HTMT; Henseler et al., 2015) y el criterio de Fornell y Larcker (1981). El ajuste global se describió con χ^2, gl, CFI, TLI, RMSEA y SRMR, interpretando este último con el punto de corte de .08 (Hu y Bentler, 1999), y se comparó con un modelo alternativo de un solo factor. La invarianza entre grupos de lengua materna se examinó con tres modelos anidados (configural, métrico y escalar) y el criterio $|\Delta CFI| < .010$ (Leitgöb et al., 2023). El

efecto indirecto se estimó como el producto a·b de los coeficientes estandarizados, con intervalos percentiles del 95 % obtenidos mediante 5 000 réplicas bootstrap con reestimación. La robustez se examinó con regresiones jerárquicas por mínimos cuadrados ordinarios con errores HC3 que incorporaron las covariables, con especificaciones alternativas (desenlace winsorizado, submuestra, predictor placebo) y con pruebas de estabilidad por jackknife, submuestreo y remuestreo Monte Carlo.

7.4 RESULTADOS

7.4.1 Modelo de medición, validez discriminante e invarianza

Las nueve cargas estandarizadas, la fiabilidad y la AVE superaron los criterios establecidos en los tres constructos (Tabla 7-2).

Tabla 7-2

Modelo de medición: propiedades psicométricas por ítem

Constructo	Ítem	λ	R^2 (SMC)	α	ω	AVE
Exposición a patrones textiles andinos	Con qué frecuencia observo o manipulo tejidos andinos en mi hogar o comunidad	0.732	0.536	0.777	0.777	0.537
	He participado en actividades de tejido, bordado o telar con familiares o en mi comunidad	0.725	0.526	0.777	0.777	0.537
	Reconozco los motivos simétricos (dualidad, espejo, recursión) presentes en los tejidos de mi entorno	0.742	0.551	0.777	0.777	0.537
Pensamiento geométrico	Identifico propiedades de figuras planas y cuerpos geométricos en situaciones cotidianas	0.656	0.430	0.778	0.778	0.540
	Resuelvo problemas de simetría, traslación y rotación de figuras	0.725	0.526	0.778	0.778	0.540
	Argumento geométricamente las soluciones a problemas de teselación y patrones	0.815	0.664	0.778	0.778	0.540

Constructo	Ítem	λ	R^2 (SMC)	α	ω	AVE
Visualización espacial	Roto mentalmente figuras tridimensionales para compararlas con un modelo	0.753	0.567	0.780	0.781	0.544
	Pliego mentalmente papel y predigo la figura resultante	0.751	0.564	0.780	0.781	0.544
	Identifico patrones espaciales ocultos en configuraciones complejas	0.707	0.500	0.780	0.781	0.544

Nota: λ = carga factorial estandarizada; R^2 (SMC) = correlación múltiple al cuadrado del ítem; α = alfa de Cronbach; ω = omega; AVE = varianza extraída promedio. Criterios: λ ≥ .50, α y ω ≥ .70, AVE ≥ .50

Los tres pares de constructos cumplieron el criterio HTMT, mientras que el criterio de Fornell-Larcker no se cumplió en la exposición ni en la visualización espacial, cuya correlación latente superó la raíz de la AVE de ambos; el pensamiento geométrico sí lo cumplió (Tabla 7-3).

Tabla 7-3

Validez discriminante

Panel A. Ratio heterotrait-monotrait (HTMT) por par de constructos			
Par de constructos	HTMT	**Criterio**	
		< .85	< .90
Exposición a patrones textiles andinos — Pensamiento geométrico	0.718	Cumple	Cumple
Exposición a patrones textiles andinos — Visualización espacial	0.752		
Pensamiento geométrico — Visualización espacial	0.667		

Panel B. Criterio de Fornell-Larcker por constructo			
Constructos	√AVE	**Correlación máxima con otro constructo**	**Cumple (√AVE > r máx.)**
Exposición a patrones textiles andinos	0.733	0.762	No
Pensamiento geométrico	0.735	0.656	Sí
Visualización espacial	0.738	0.762	No

Nota: HTMT calculada sobre los ítems del modelo de medición (Henseler et al., 2015); criterios < .85 (estricto) y < .90 (liberal). Fornell-Larcker: se cumple cuando la raíz de la AVE del constructo supera su mayor correlación con otro constructo (Fornell y Larcker, 1981)

El modelo estructural hipotetizado reprodujo la matriz observada con un ajuste excelente, en tanto que el modelo alternativo de un factor se ajustó claramente peor (Tabla 7-4). El SRMR del modelo de medición fue de .033 y el factor de inflación de la varianza máximo fue de 1.013. Las diferencias de CFI entre los niveles configural, métrico y escalar de invarianza por lengua materna se mantuvieron por debajo del criterio de .010.

Tabla 7-4

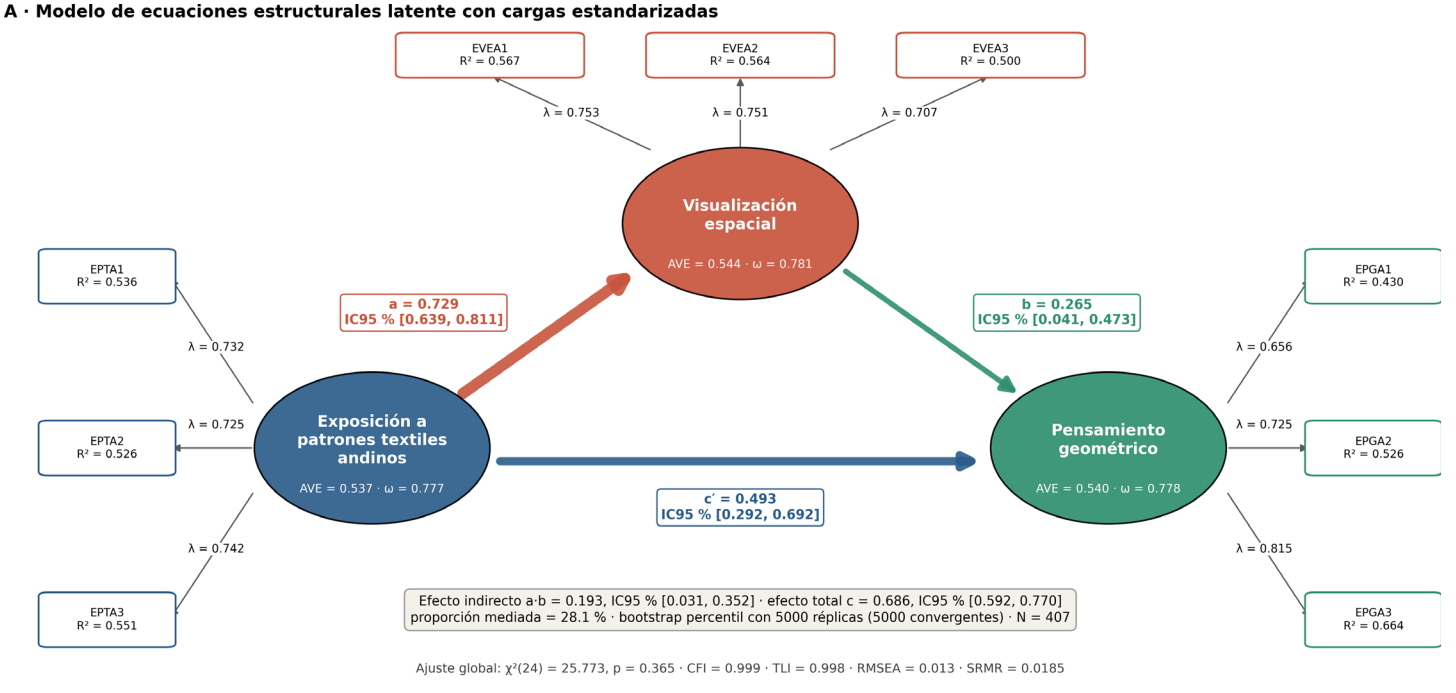
Ajuste global del modelo e invarianza de medición por lengua materna

Modelo		χ^2	gl	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA
Modelo	estructural hipotetizado (3 factores)	25.773	24	0.999	0.998	0.013	0.0185	—	—
	alternativo (1 factor)	202.01	27	0.872	0.830	0.126	0.097	—	—
Invarianza por lengua materna:	configural	—	—	0.991	0.986	0.026	—	—	—
	métrica	—	—	0.992	0.988	0.034	—	0.001	0.008
	escalar	—	—	0.992	0.988	0.034	—	0.000	0.001

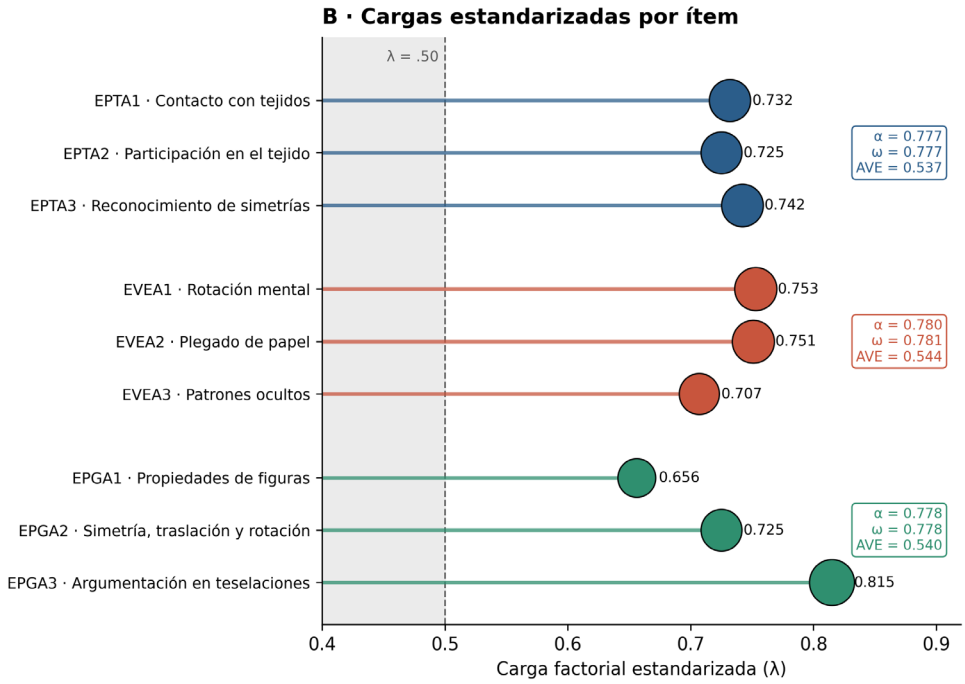
Nota: El modelo alternativo carga los nueve ítems en un único factor. Invarianza: el modelo configural se estima por separado en cada grupo de lengua materna (χ^2 y gl agregados); el métrico, con parámetros comunes sobre la muestra combinada; y el escalar, sobre los ítems centrados en la media de su grupo. ΔCFI y ΔRMSEA se calculan respecto del nivel anterior; criterio |ΔCFI| < .010 (Leitgöb et al., 2023). SRMR: punto de corte .08 (Hu y Bentler, 1999).

Figura 7-1

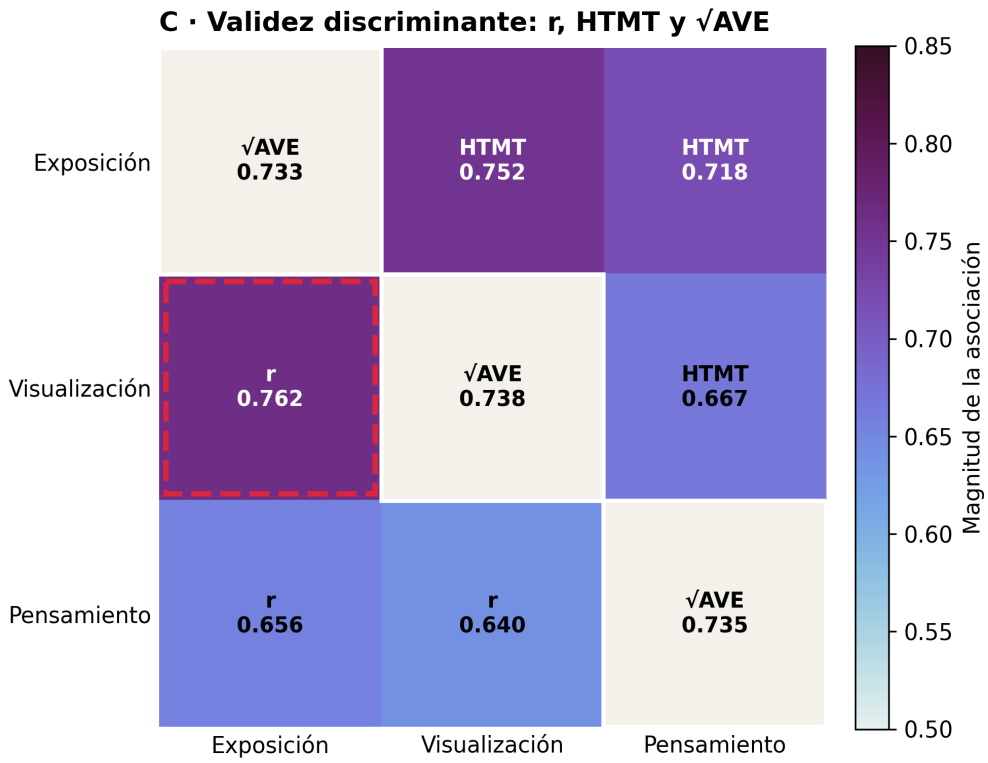
Modelo de medida y estructural de la exposición a patrones textiles andinos, la visualización espacial y el pensamiento geométrico



Nota. N = 407; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel A: constructos latentes (elipses) con sus tres indicadores (rectángulos), cargas estandarizadas (λ = 0.656–0.815), R^2 de cada ítem y rutas estructurales con su intervalo de confianza del 95 % bootstrap percentil (5000 réplicas): a = 0.729 [0.639, 0.811], b = 0.265 [0.041, 0.473] y c′ = 0.493 [0.292, 0.692]; el grosor de cada flecha es proporcional a |β|. El efecto indirecto es a·b = 0.193 [0.031, 0.352] y el total c = 0.686 [0.592, 0.770] (28.1 % mediado); ajuste: $\chi^2(24) = 25.773$, p = .365, CFI = 0.999, TLI = 0.998, RMSEA = 0.013, SRMR = 0.0185.



Nota. N = 407; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel B: carga estandarizada de cada ítem; el tamaño del punto es proporcional a su R^2 (0.430–0.664), la línea discontinua marca λ = .50 y los recuadros muestran α, ω y AVE de cada constructo (AVE = 0.537–0.544).



Nota. N = 407; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel C: matriz de validez discriminante con la correlación entre constructos en el triángulo inferior, la HTMT en el superior y la √AVE en la diagonal; el borde discontinuo señala el par exposición–visualización espacial, cuya correlación (r = 0.762) supera la √AVE de ambos constructos (0.733 y 0.738), mientras que todas las HTMT son inferiores a .85 (máximo 0.752).

La Figura 7-1 sintetiza el modelo de medida y estructural, las propiedades psicométricas de los ítems y la matriz de validez discriminante.

7.4.2 Descriptivos y modelo estructural

Las puntuaciones de los tres constructos presentaron distribuciones aproximadamente simétricas y correlaciones positivas y elevadas entre sí, mientras que sus correlaciones con las covariables fueron bajas (Tabla 7-5).

Tabla 7-5

Estadísticos descriptivos y correlaciones

	Variable	M	DE	Asimetría	Curtosis	1	2	3	4
1	Exposición a patrones textiles andinos	2.999	0.661	−0.028	0.170	—	0.656*** [0.597, 0.708]	0.762*** [0.718, 0.800]	0.059 [−0.038, 0.155]
2	Pensamiento geométrico	2.999	0.664	−0.026	0.022	0.656 ***	—	0.640*** [0.579, 0.694]	0.090 [−0.008, 0.185]
3	Visualización espacial	2.999	0.659	0.121	0.467	0.762 ***	0.640***	—	0.063 [−0.034, 0.159]
4	Memoria de trabajo visoespacial	4.393	3.132	1.393	2.233	0.059	0.090	0.063	—
5	Lengua materna	0.506	0.501	−0.025	−1.999	−0.066	−0.048	−0.071	0.001
6	Escolaridad discontinua	0.472	0.500	0.113	−1.987	0.064	0.099*	0.071	0.062
7	Sexo	0.477	0.500	0.093	−1.991	−0.035	−0.081	0.009	0.067

Nota: N = 407. Triángulo superior: r con IC95 % (transformación Z de Fisher) entre corchetes; triángulo inferior: r. Los constructos se expresan en la escala Likert de 1 a 5; las covariables dicotómicas, como proporción del valor 1. * p < .05; *** p < .001.

En el modelo estructural, las tres rutas hipotetizadas resultaron positivas y significativas, y el efecto indirecto alcanzó significación (Tabla 7-6).

Tabla 7-6

Rutas estructurales del modelo CB-SEM

Ruta	β estandarizado	EE	p	Hipótesis
Exposición a patrones textiles andinos → Pensamiento geométrico	0.493	0.090	< .001	H1 (c′)
Visualización espacial → Pensamiento geométrico	0.265	0.086	.01	H3 (b)
Exposición a patrones textiles andinos → Visualización espacial	0.729	0.073	< .001	H2 (a)
Exposición a patrones textiles andinos → Visualización espacial → Pensamiento geométrico (indirecto)	0.193	—	.024	H4 (a·b)

Nota: Coeficientes estandarizados del modelo estructural latente estimado por máxima verosimilitud. El efecto indirecto se contrasta con su intervalo bootstrap en la Tabla 7-7. Ajuste: CFI = .999; TLI = .998; RMSEA = .013; SRMR = .0185.

Tabla 7-7

Descomposición de los efectos (mediación)

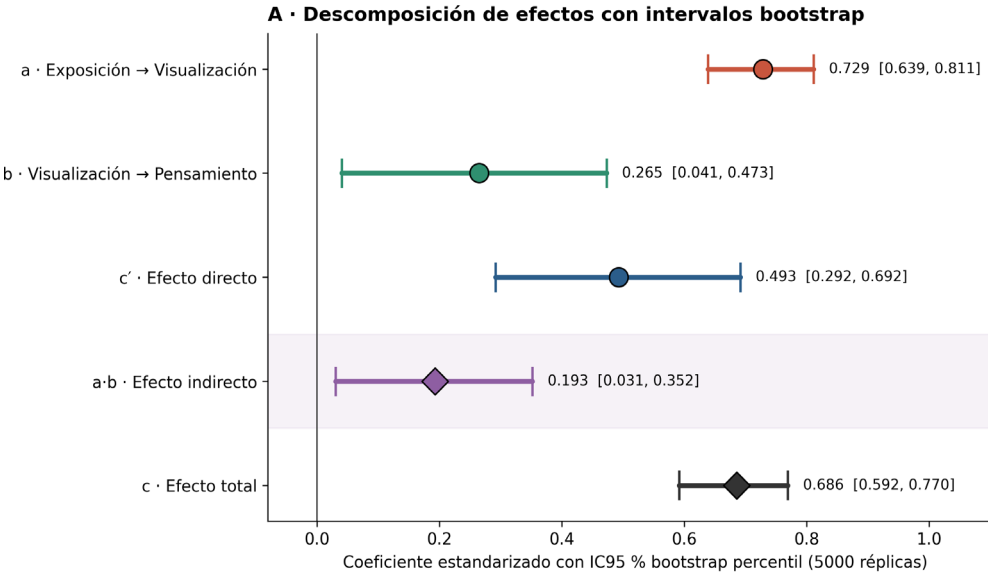
Efecto		Coeficiente	IC95 %		Significativo
			inferior	superior	
a:	Exposición → Visualización espacial	0.729	0.639	0.811	Sí
b:	Visualización espacial → Pensamiento geométrico	0.265	0.041	0.473	Sí
c′:	efecto directo	0.493	0.292	0.692	Sí
a·b:	efecto indirecto	0.193	0.031	0.352	Sí
c:	efecto total (c′ + a·b)	0.686	0.592	0.770	Sí

Nota: Coeficientes estandarizados del modelo latente. Intervalos de confianza percentiles del 95 % con 5 000 réplicas bootstrap (5 000 convergentes). Un efecto es significativo cuando su intervalo excluye el cero.

La descomposición de efectos mostró que el efecto indirecto y el directo excluyeron el cero en sus intervalos, configurando una mediación parcial en la que la vía a través de la visualización espacial transmitió el 28.1 % del efecto total (Tabla 7-7).

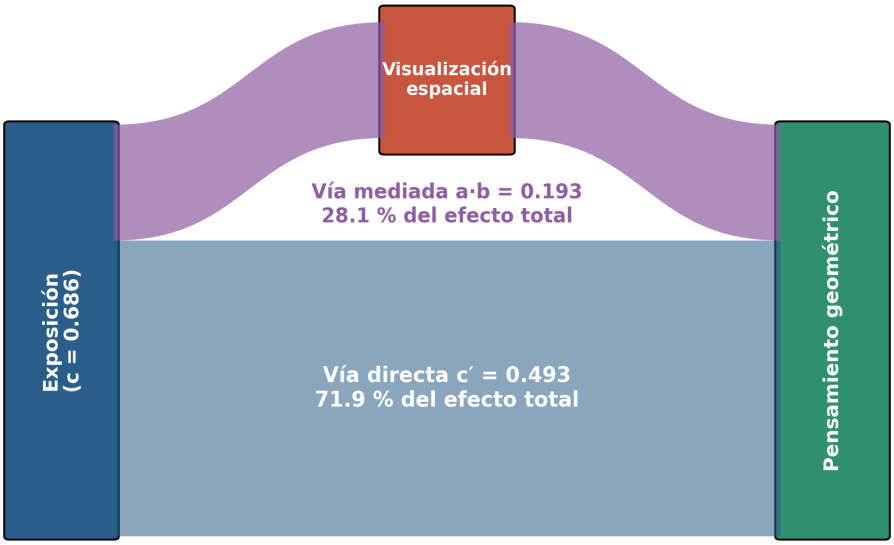
Figura 7-2

Descomposición del efecto de la exposición a patrones textiles andinos sobre el pensamiento geométrico

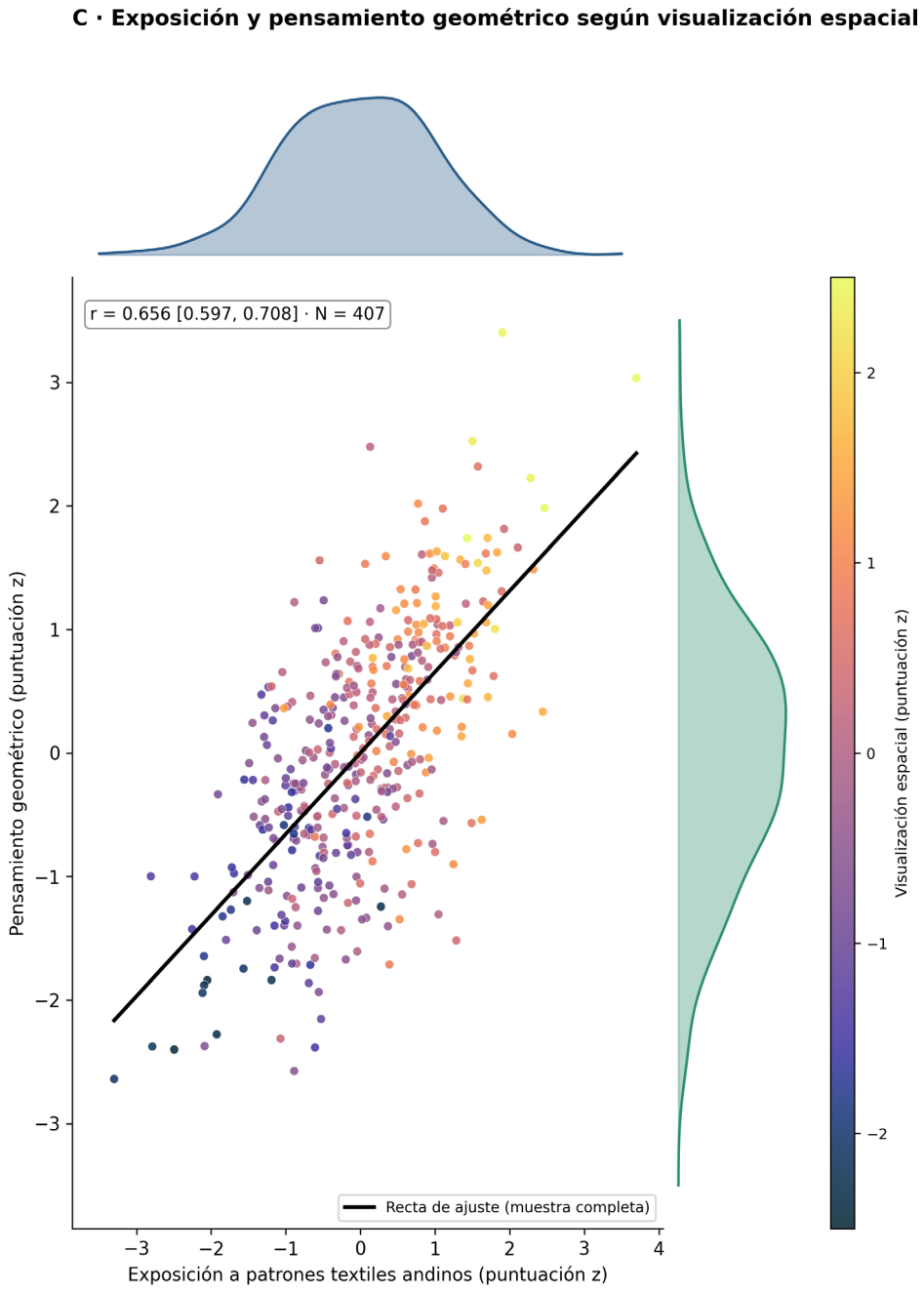


Nota. N = 407. Panel A: coeficientes estandarizados del modelo latente con intervalos de confianza del 95 % bootstrap percentil (5000 réplicas) para las rutas a (0.729 [0.639, 0.811]), b (0.265 [0.041, 0.473]) y c′ (0.493 [0.292, 0.692]), el efecto indirecto a·b (0.193 [0.031, 0.352]) y el efecto total c (0.686 [0.592, 0.770]); ningún intervalo incluye el cero.

B · Flujo del efecto total: vía directa y vía mediada



Nota. N = 407. Panel B: diagrama de flujo del efecto total, con el ancho de cada banda proporcional a su contribución: la vía directa aporta el 71.9 % (c′ = 0.493) y la vía mediada por la visualización espacial el 28.1 % (a·b = 0.193).



Nota. N = 407.Panel C: dispersión del pensamiento geométrico frente a la exposición en puntuaciones z de los compuestos, con el color de cada punto según la visualización espacial, recta de ajuste de la muestra completa y densidades marginales; r = 0.656 [0.597, 0.708]. Los estudiantes con mayor exposición se concentran en valores altos de visualización espacial y de pensamiento geométrico.

La Figura 7-2 representa la descomposición del efecto total y la relación entre exposición y pensamiento geométrico según el nivel de visualización espacial. Las hipótesis H1, H2, H3 y H4 recibieron soporte empírico en el modelo latente.

7.4.3 Robustez

Las regresiones jerárquicas con las covariables reprodujeron el patrón del modelo latente: la exposición aportó un incremento sustancial de varianza explicada, su coeficiente se redujo al incorporar la visualización espacial, y entre las covariables solo el sexo alcanzó significación en el modelo completo (Tabla 7-8).

Tabla 7-8

Robustez: regresión jerárquica por mínimos cuadrados ordinarios con covariables

Predictor	M1: covariables	M2: + exposición	M3: + visualización espacial
Memoria de trabajo visoespacial	0.020 [−0.000, 0.040]	0.011 [−0.004, 0.027]	0.010 [−0.005, 0.026]
Lengua materna	−0.062 [−0.195, 0.071]	−0.005 [−0.106, 0.096]	0.005 [−0.093, 0.102]
Escolaridad discontinua	0.135* [0.000, 0.269]	0.079 [−0.022, 0.181]	0.070 [−0.028, 0.169]
Sexo	−0.125 [−0.259, 0.008]	−0.089 [−0.191, 0.013]	−0.105* [−0.203, −0.007]
Exposición a patrones textiles andinos	—	0.444*** [0.395, 0.494]	0.269*** [0.193, 0.344]
Visualización espacial	—	—	0.347*** [0.240, 0.455]
R^2	0.027	0.440	0.487
ΔR^2	—	0.413	0.047
F (p)	2.893 (.022)	63.292 (< .001)	72.319 (< .001)

Nota: Desenlace: pensamiento geométrico. Coeficientes de regresión con IC95 % basados en errores estándar HC3. * p < .05; *** p < .001. Al no declararse moderadores, el modelo con términos de interacción reproduce exactamente el M3 (R^2 = .487) y no se presenta como columna adicional. Estas regresiones se estiman sobre puntuaciones compuestas y constituyen una comprobación de robustez del modelo latente, no su estimación principal.

El efecto de la exposición se mantuvo positivo y significativo con el desenlace winsorizado, en la submuestra del 60 % y sin términos de interacción, mientras que el predictor placebo no mostró asociación (Tabla 7-9).

Tabla 7-9

Robustez: especificaciones alternativas)

Especificación	n	b (exposición)	EE	IC 95 %	t	R^2
Desenlace winsorizado al 5 %	407	0.395	0.022	[0.353, 0.438]	< .001	0.417
Submuestra del 60 %	244	0.439	0.034	[0.373, 0.505]	< .001	0.420
Especificación sin términos de interacción	407	0.444	0.025	[0.395, 0.494]	< .001	0.440
Predictor placebo (exposición permutada)	407	0.016	0.037	[−0.057, 0.088]	.669	0.028

Nota: Regresiones por mínimos cuadrados ordinarios del pensamiento geométrico sobre la exposición y las covariables, sin el mediador. El predictor placebo reasigna al azar los valores de la exposición entre los estudiantes; su efecto nulo indica que la asociación observada no es un artefacto de la especificación.

Tabla 7-10

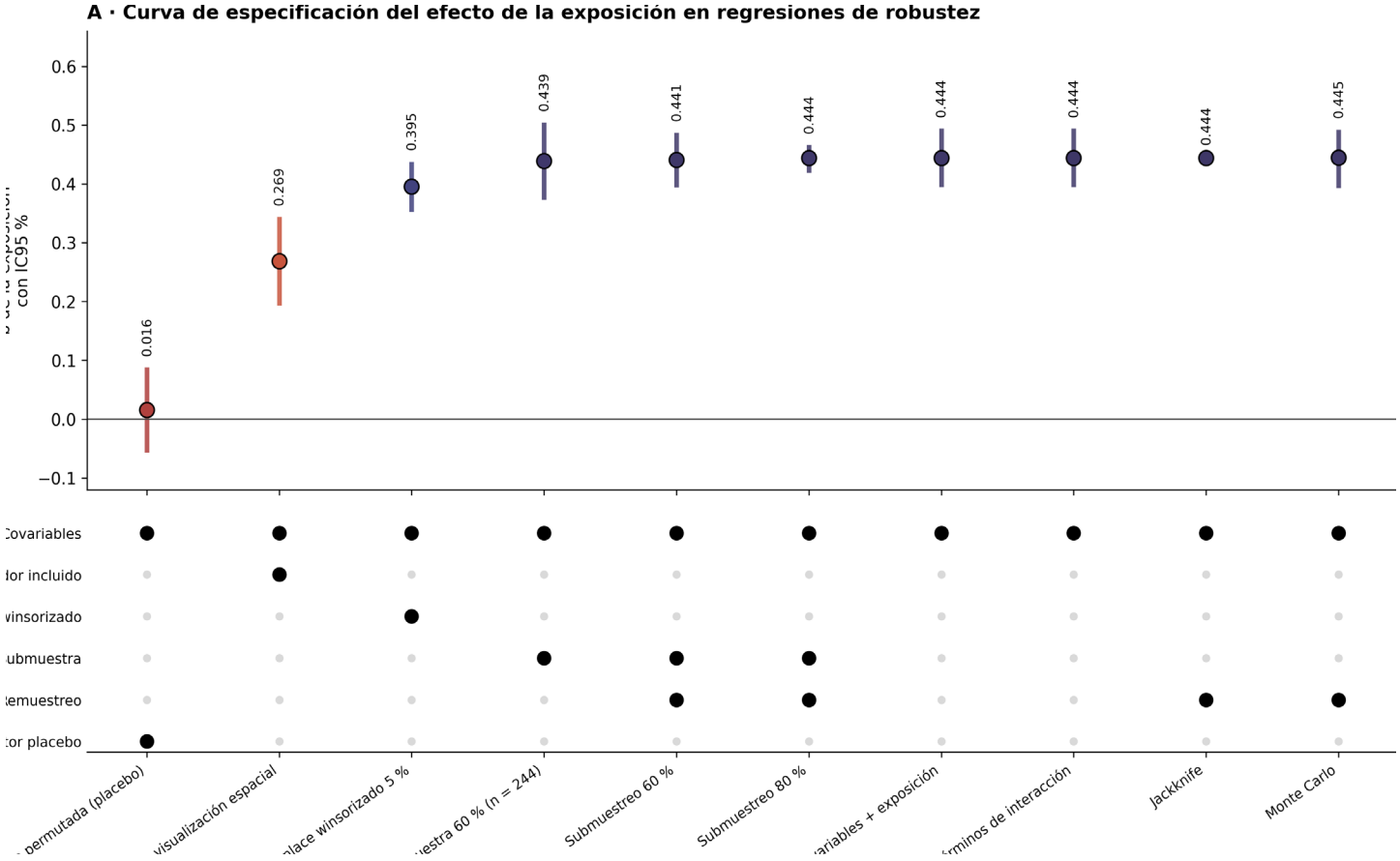
Robustez: estabilidad del efecto de la exposición ante perturbaciones de la muestra

Prueba		Réplicas	β medio	DE	iIC 95 %	Signo estable (%)	Significativo (%)
Jackknife (exclusión de un caso)		407	0.4444	0.0013	[0.4418, 0.4473]	100 %	100 %
Submuestreo del...	80 %	100	0.4439	0.0141	[0.4194, 0.4663]	100 %	100 %
	60 %	100	0.4411	0.0218	[0.3945, 0.4872]	100 %	100 %
Remuestreo Monte Carlo (bootstrap)		1000	0.4447	0.0246	[0.3928, 0.4923]	100 %	100 %

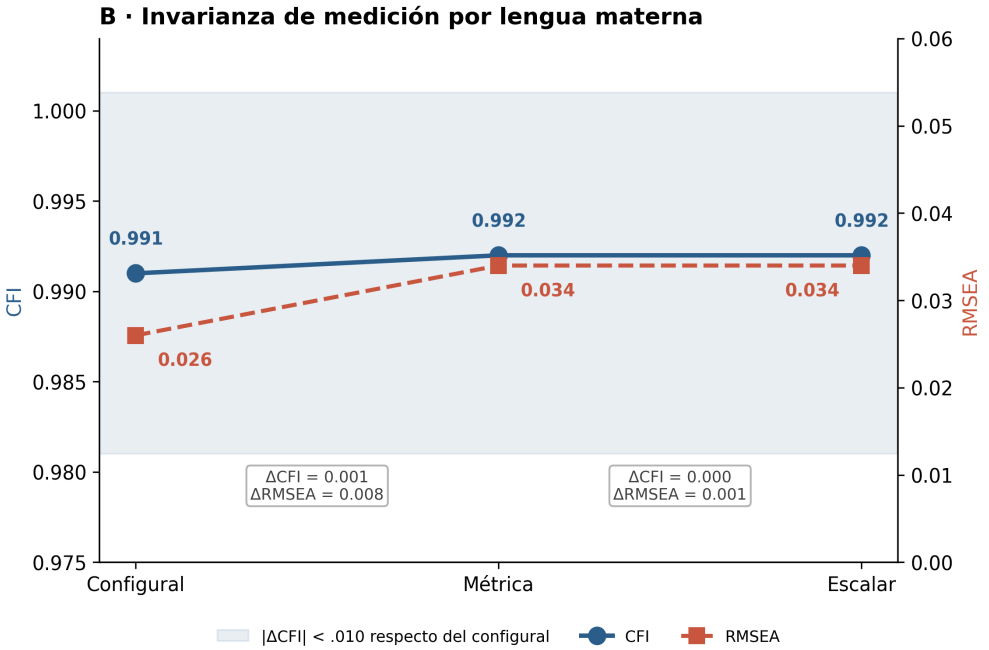
Nota: Coeficiente de la exposición en la regresión del pensamiento geométrico sin el mediador. Signo estable: porcentaje de réplicas con el mismo signo que la estimación completa; significativo: porcentaje con p < .0.

Figura 7-3

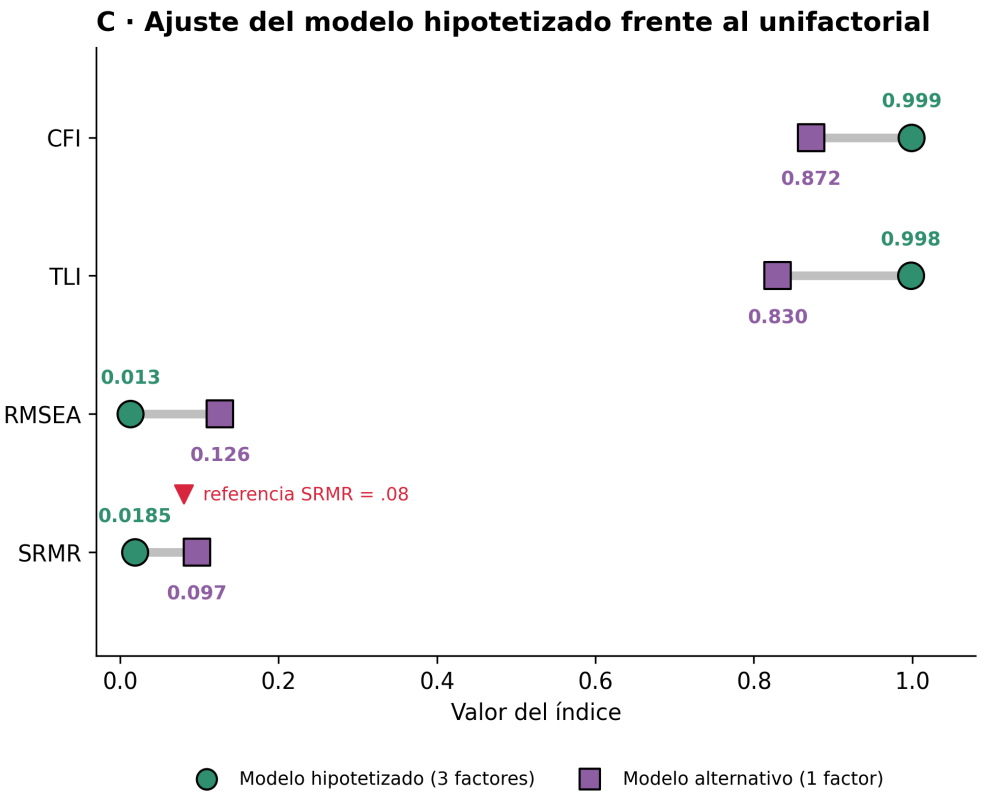
Robustez del efecto de la exposición, invarianza de medición y ajuste comparado del modelo



Nota. N = 407. Panel A: curva de especificación del coeficiente de la exposición (b, con IC95 %) en las regresiones de robustez por mínimos cuadrados ordinarios, ordenadas de menor a mayor, con la matriz inferior que indica qué rasgos incluye cada especificación; el coeficiente se mantiene entre 0.395 y 0.445 sin el mediador, desciende a 0.269 [0.193, 0.344] al incorporar la visualización espacial y se anula con el predictor placebo (0.016 [−0.057, 0.088]).



Nota. N = 407. Panel C: índices de ajuste del modelo hipotetizado de tres factores (CFI = 0.999, TLI = 0.998, RMSEA = 0.013, SRMR = 0.0185) frente al modelo de un factor (CFI = 0.872, TLI = 0.830, RMSEA = 0.126, SRMR = 0.097); el triángulo rojo sobre la fila del SRMR marca la referencia SRMR = .08 (Hu y Bentler, 1999). Los coeficientes del Panel A proceden de regresiones sobre puntuaciones compuestas y no son comparables en escala con los coeficientes estandarizados del modelo latente.



Nota. N = 407. Panel C: índices de ajuste del modelo hipotetizado de tres factores (CFI = 0.999, TLI = 0.998, RMSEA = 0.013, SRMR = 0.0185) frente al modelo de un factor (CFI = 0.872, TLI = 0.830, RMSEA = 0.126, SRMR = 0.097); el triángulo rojo sobre la fila del SRMR marca la referencia SRMR = .08 (Hu y Bentler, 1999). Los coeficientes del Panel A proceden de regresiones sobre puntuaciones compuestas y no son comparables en escala con los coeficientes estandarizados del modelo latente.

Por último, el coeficiente de la exposición conservó su signo y su significación en la totalidad de las réplicas de jackknife, submuestreo y remuestreo Monte Carlo (Tabla 7-10).

La Figura 7-3 reúne la curva de especificación, la invarianza de medición y el ajuste comparado de los modelos.

7.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre exposición y pensamiento geométrico, sino que la ruta más intensa del modelo une la exposición con la visualización espacial (H2; β = .729), por encima de la vía directa (H1; β = .493), hasta rozar el límite de lo empíricamente separable: la correlación latente entre ambos constructos supera la raíz de la AVE de los dos, aunque la HTMT se mantiene por debajo de los umbrales usuales. Surge así una pregunta incómoda: ¿mide la escala de visualización espacial una capacidad distinta o refleja, en clave de autopercepción, la misma familiaridad con el textil que recoge la escala de exposición? Dos evidencias disuaden de una lectura reduccionista: el modelo de un factor se ajusta mucho peor que el de tres, y la invarianza métrica y escalar por lengua materna indica que los ítems funcionan de forma equivalente en ambos grupos. Aun así, la magnitud de a debe leerse con cautela, y el patrón sugiere, sin demostrarlo, que quienes conviven con los tejidos se perciben también más hábiles para transformar mentalmente las formas.

Desde una perspectiva crítica, la mediación hallada es modesta. El efecto indirecto transmite el 28.1 % del efecto total, el límite inferior de su intervalo se aproxima a cero y el intervalo de la ruta b es amplio, de modo que la mayor parte de la asociación viaja por la vía directa. Esa preponderancia es consistente con una lectura menos cognitivista del textil: el artefacto no solo ofrece configuraciones que rotar o reflejar, sino que encierra decisiones técnicas culturalmente situadas, como las que organizan la construcción de textiles de urdimbre y trama discontinuas (Phipps, 2025), y dimensiones materiales, rituales y energéticas del quehacer textil que el modelo no

captura (Berduccy, 2026). Algo semejante ocurre con otros soportes andinos: el sistema simbólico pasto, presente en el tejido, la agricultura y la organización territorial, expresa el conocimiento matemático de forma contextualizada y ligada a la experiencia cotidiana (Sono et al., 2026); las familias altoandinas emplean unidades de medida corporales, como los dos pies para sembrar la papa en el surco «iskay chaki» (Hinojosa et al., 2024); y los saberes matemáticos de la comunidad asháninca aparecen aplicados de forma empírica y antropométrica en artesanías, juegos y ceremonias (Salvatierra, 2026). Si el conocimiento geométrico se transmite en estos entornos mediante la medición con el cuerpo, la narración y el rito, la visualización espacial autopercibida capturaría solo una fracción del mecanismo, y la parte directa del efecto podría albergar vías no modeladas, como el reconocimiento de la simetría como estructura, la motivación o la identidad cultural. La naturaleza de autoinforme de las tres escalas añade una explicación competidora: parte de la covariación podría deberse a una disposición general a valorarse positivamente en todos los dominios, una posibilidad que el diseño transversal no permite descartar. La solidez del efecto de la exposición ante las especificaciones alternativas, las pruebas de estabilidad y el predictor placebo descarta que sea un artefacto de la especificación, pero no resuelve esta ambigüedad de medida.

En el plano de la coincidencia con la teoría, la relación entre visualización espacial y pensamiento geométrico (H3; $\beta = .265$) concuerda con la teoría de van Hiele (1986), según la cual el reconocimiento de transformaciones sostiene el paso hacia el análisis de propiedades, y con la evidencia metaanalítica de que el entrenamiento espacial transfiere al rendimiento matemático (Hawes et al., 2022). La asociación entre exposición y visualización (H2) es coherente con la maleabilidad de las habilidades espaciales: decodificar un textil exige rotar, reflejar y descomponer motivos. Que el desarrollo del pensamiento geométrico de estudiantes andinos peruanos pueda describirse con los niveles de van Hiele, como en la enseñanza de cuadriláteros en una zona rural de Huancavelica (Huamán y Carhuapoma, 2024), ofrece un punto de anclaje adicional para esta lectura.

La contribución teórica reside en articular tres marcos que por separado no anticipan la mediación parcial observada: la etnomatemática identifica el artefacto como sistema de conocimiento, la cognición espacial aporta el mecanismo y los niveles de van Hiele trazan el puente hacia el pensamiento geométrico formal. Este patrón precisa la propuesta etnomatemática (D'Ambrosio, 1985): el artefacto cultural no actúa solo como contenido o como contexto motivador, sino como un entorno de práctica de transformaciones mentales que el aprendizaje geométrico escolar puede capitalizar. El modelo es aplicable más allá del caso andino, a cualquier sistema educativo cuyos estudiantes convivan con un patrimonio artesanal de estructura geométrica, y la invarianza entre estudiantes con lengua materna originaria y con castellano indica que la medición puede emplearse para comparar grupos lingüísticos. Esta equivalencia cobra relevancia porque la contextualización lingüística y cultural de la enseñanza se asocia con el aprendizaje matemático de niños quechuahablantes (Saldivar et al., 2025); el presente resultado no evalúa esa enseñanza, pero aporta la condición de comparabilidad métrica que los estudios sobre ella requieren.

Las implicaciones prácticas se dirigen a la enseñanza de la geometría en contextos con patrimonio textil. En primer lugar, los textiles pueden incorporarse a las unidades de simetría, transformaciones y teselación como material de práctica de rotación y reflexión, no como ilustración decorativa, en línea con la propuesta de la etnomatemática como estrategia para elevar los niveles de logro en geometría en la educación básica peruana (Ramos y Ramos, 2023) y con la demanda de incorporar elementos propios del entorno para lograr un aprendizaje significativo en las escuelas andinas (Mallco, 2025). Otros artefactos, como la yupana o ábaco inca, ofrecen posibilidades análogas de aplicación en la educación matemática latinoamericana (Vilca et al., 2023), y símbolos geométricos como la chakana y sus colores han sido propuestos como recurso didáctico en la educación intercultural bilingüe (Mayanza y Mora, 2022). Desde la descomposición geométrica explicada desde la cosmovisión andina hasta

la enseñanza del álgebra (Turriaga y Sgreccia, 2023), la transferencia de saberes andinos al aula parece factible en distintos dominios.

En segundo lugar, dado que una parte del vínculo se transmite por la visualización espacial, las tareas deberían exigir explícitamente la transformación mental de motivos reales: anticipar el resultado de un giro, completar una reflexión, descomponer una teselación. Los entornos digitales pueden complementar esa práctica; una propuesta con GeoGebra fortaleció capacidades en transformaciones geométricas en estudiantes de secundaria de Ayabaca (Ruiz et al., 2026), y su uso se vincula con el logro de competencias matemáticas en la secundaria rural de Puno (Ruelas y Mamani, 2026), aunque estos estudios no incorporan el patrimonio textil, y combinar ambos recursos es una hipótesis por contrastar. En tercer lugar, la formación docente debería incluir la lectura geométrica de los artefactos locales, además de su valor cultural, pues las decisiones del profesor al elegir estrategias y recursos durante la planificación condicionan la enseñanza de la geometría (Advíncula et al., 2024).

7.6 CONCLUSIONES

La exposición a patrones textiles andinos se asocia con el pensamiento geométrico de estudiantes de secundaria del Perú directamente y a través de la visualización espacial, en un patrón de mediación parcial en el que la vía indirecta transmite poco más de una cuarta parte del efecto total. Las cuatro hipótesis recibieron soporte en un modelo de ecuaciones estructurales de ajuste excelente, con invarianza métrica y escalar entre estudiantes de lengua materna originaria y castellana. El conocimiento geométrico codificado en un artefacto cultural se vincula con el aprendizaje escolar en la medida en que su decodificación pone en juego la transformación mental de formas, y este mecanismo es transferible a otros patrimonios artesanales con estructura geométrica.

Este estudio presenta limitaciones que acotan sus conclusiones. El diseño transversal no permite establecer el orden temporal de las relaciones ni afirmar causalidad en la mediación. La muestra se obtuvo por

conveniencia, por lo que los resultados describen al grupo analizado y no se generalizan a la población estudiantil peruana. Los tres constructos se midieron con escalas de autoinforme, que reflejan la percepción de los estudiantes y no su desempeño en pruebas de ejecución. La agrupación de los estudiantes en aulas e instituciones no se modeló, lo que puede subestimar los errores estándar. Finalmente, la proximidad empírica entre exposición y visualización espacial, patente en el incumplimiento del criterio de Fornell-Larcker, limita la separación de ambos constructos.

La investigación futura debería contrastar la mediación con diseños longitudinales o con intervenciones basadas en textiles, sustituir el autoinforme por pruebas de ejecución de rotación mental y razonamiento geométrico, modelar la estructura de aulas e instituciones con modelos multinivel e identificar qué rasgos del artefacto (tipo de simetría, escala, complejidad del motivo) se vinculan más estrechamente con la visualización espacial. Asimismo, convendría explorar las vías no espaciales de la asociación directa, de naturaleza simbólica, técnica y cultural, para completar el cuadro del mecanismo que este trabajo apenas delimita.

7.7 **REFERENCIAS BIBLIOGRÁFICAS**

Advíncula, E., Torres, I., Carreño, E., Hau, F., & Montes, M. (2024). Conocimiento especializado de un profesor de educación secundaria al diseñar clases de cuadriláteros. *PNA. Revista de Investigación en Didáctica de la Matemática, 18*(4), 415-438. https://doi.org/10.30827/pna.v18i4.27637

Atupaña, J., Cunduri, M., Patarón, N., Yunda, J., & Pucha, M. (2025). Enseñanza de la etnomátematica a través de la confección del maki poncho en la cultura puruha. *Ciencia Latina Revista Científica Multidisciplinar, 9*(2), 2896-2915. https://doi.org/10.37811/cl_rcm.v9i2.17101

Berduccy, S. (2026). The Electric Thread: Andean Textile Practices and the Energy That Activates Them. *Leonardo*, *59*(3), 248-255. https://doi.org/10.1162/leon.a.2680

Bonilla, M. (2023). Ethnomathematics and complexity: A study of the process of elaboration of a Peruvian Andean textile. In M. Borba & D. Orey (Eds.), *Ubiratan D'Ambrosio and mathematics education* (pp. 179-200). Springer. https://doi.org/10.1007/978-3-031-31293-9_13

D'Ambrosio, U. (1985). Ethnomathematics and its place in the history and pedagogy of mathematics. *For the Learning of Mathematics*, *5*(1), 44-48.

Davila, J. (2026). Incidencia de la educación artística en el pensamiento abstracto-matemático de estudiantes de secundaria de Lima. *Revista Simón Rodríguez*, *6*(12), 1-14. https://doi.org/10.62319/simonrodriguez.v.6i12.204

Flores, K., & Otáñez, J. (2025). Estudio de los elementos gráficos en la indumentaria del Cacique Tisaleo, personaje principal de la festividad del Inga Palla. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, *6*(1). https://doi.org/10.56712/latam.v6i1.3552

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Fossum, J., & Montoya, A. (2023). When to use different inferential methods for power analysis and data analysis for between-subjects

mediation. *Advances in Methods and Practices in Psychological Science*, *6*(2). https://doi.org/10.1177/25152459231156606

Gilligan, K., Hawes, Z., & Mix, K. (2022). Spatial thinking as the missing piece in mathematics curricula. *npj Science of Learning*, *7*(1). https://doi.org/10.1038/s41539-022-00128-9

Gilsdorf, T. (2024). Género, matemáticas culturales, y arte textil de los Mazahua y Hñähñu. *Revista Latinoamericana de Etnomatemática Perspectivas Socioculturales de la Educación Matemática*, *16*, 1-23. https://doi.org/10.22267/relatem.22152.91

Hartoyo, A., Fitriawan, D., Siregar, N., & Ganda, F. (2025). Connecting cultural roots with mathematical thinking: A comprehensive meta-analysis of ethnomathematics practices in indonesian classrooms. *Al-Jabar : Jurnal Pendidikan Matematika*, *16*(2), 673-689. https://doi.org/10.24042/ajpm.v16i2.28734

Hawes, Z., Gilligan, K., & Mix, K. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, *58*(1), 112-137. https://doi.org/10.1037/dev0001281

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hinojosa, J., Yucra, M., Callomamani, J., Enríquez, P., Mamani, J., Catacora, E., Jilaja, E., & Poma, H. (2024). Praxis de la etnomatemática quechua en la región alto andina del Perú. *Inclusão Digital no Ensino*

de Ciências e Matemática: um olhar para a diversidade, 56-79. https://doi.org/10.37885/240115596

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huamán, A., & Carhuapoma, L. (2024). Modelo de Van Hiele en el aprendizaje de cuadriláteros, en estudiantes del cuarto grado de educación secundaria, Huancavelica. *Quintaesencia, 15*(1), 24-32. https://doi.org/10.54943/rq.v15i.556

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research, 110,* 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Maldonado, C., Aguirre, C., del Río, M., & Álvarez, J. (2026). Puruhá Symbols on Guano Rugs: A Semiotic Approach to Cultural Continuity. *Heritage*, *9*(5), 167. https://doi.org/10.3390/heritage9050167

Mallco, J. (2025). La importancia de la etnomatemática en las comunidades andinas. *Llimpi*, *5*(2), 17-23. https://doi.org/10.54943/lree.v5i2.617

Mayanza, L., & Mora, A. (2022). Sabiduría andina chakana y sus colores, una herramienta didáctica para la educación intercultural bilingüe. *Diálogo andino*(67), 99-111. https://doi.org/10.4067/s0719-26812022000100099

Mendoza, L., Dudojć, O., & Jordan, F. (2026). Maths and crafts: embodied mathematical knowledge in fibre technologies. *Language Sciences, 119*, 101874. https://doi.org/10.1016/j.langsci.2026.101874

Merma, D., Oscco, R., Huaman, J., Ccorimanya, E., & Pacheco, M. (2024). *Visualización espacial y razonamiento geométrico mediante estrategias didácticas: Modelo Van Hiele y Geogebra.* Arco Editores. https://doi.org/10.48209/978-65-5417-228-8

Morales, J. (2024). Saberes Matemáticos Ancestrales: una mirada Etnomatemática al tejido en lana de ovejo. *Revista Latinoamericana de Etnomatemática Perspectivas Socioculturales de la Educación Matemática, 16*, 57-77. https://doi.org/10.22267/relatem.22152.94

Palomino, P., Napurí, A., & Heckman, A. (2024). El tejido como expresión integral del mundo de la mujer acchahuata en la región de Cusco, Perú. *Diálogo andino*(75), 187-200. https://doi.org/10.4067/s0719-26812024000300187

Park, H. (2022). A Study on Geometric Thinking in Andean Ethnophilosophy. *The Journal of Humanities and Social sciences 21, 13*(5), 2777-2792. https://doi.org/10.22143/hss21.13.5.194

Phipps, E. (2025). Andean Traditions in the Making of Discontinuous Warp (and Weft) Textiles. *The Textile Museum Journal, 52*(1), 24-43. https://doi.org/10.1353/tmj.00013

Ramos, P., & Ramos, R. (2023). *La etnomatemática como estrategia de aprendizaje en la geometría.* Fondo Editorial Professionals On Line. https://doi.org/10.47422/fepol.15

Rosa, M., & Orey, D. (2025). Reflexões Internacionais sobre as Características Decolonizadoras do Programa Etnomatemática. *PARADIGMA*, e2025014. https://doi.org/10.37618/paradigma.1011-2251.2025.e2025014.id1604

Ruelas, E., & Mamani, B. (2026). El uso del software GeoGebra y el desempeño por competencias matemáticas en la educación secundaria rural. *Technological Innovations Journal, 5*(3), 7-20. https://doi.org/10.35622/j.ti.2026.03.001

Ruiz, N., Armas, R., Correa, R., Fiestas, J., & Balcázar, J. (2026). GeoGebra para fortalecer capacidades matemáticas en transformaciones geométricas en estudiantes de secundaria de Ayabaca. *Revista Tribunal, 6*(15), 271-285. https://doi.org/10.59659/revistatribunal.v6i15.398

Saldivar, L., Rojas, H., & Yanqui, F. (2025). Implementation of a Digital Educational Tool in the Quechua Language for Learning Mathematics among Quechua-Speaking Children. *International Journal of Learning, Teaching and Educational Research, 24*(2), 352-374. https://doi.org/10.26803/ijlter.24.2.18

Salvatierra, Á. (2026). Vivencias y conceptos matemáticos: etnomatemática en la comunidad asháninca. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, *24*(2), 1-27. https://doi.org/10.11600/rlcsnj.24.2.7622

Sono, D., Palacios, F., Aroca, A., Padilla, S., & Yandún, C. (2026). Simbología Pasto, movimiento corporal y aprendizaje matemático en educación intercultural: una revisión sistemática. *Retos*, *79*, 616-633. https://doi.org/10.47197/retos.v79.119023

Turriaga, M., & Sgreccia, N. (2023). Integración de la descomposición geométrica y la cosmovisión andina en la enseñanza del álgebra en 9no EGB en Ecuador. *Revemop*, *5*, e202304. https://doi.org/10.33532/revemop.e202304

Valle, L. (2022). La herencia gráfica de los tocapus en el pallay de los tejidos de Chinchero. *Escritura y Pensamiento*, *21*(43), 143-153. https://doi.org/10.15381/escrypensam.v21i43.21873

Vilca, H., Mamani, W., Maraza, B., & Bizarro, W. (2023). Yupana o ábaco inca, a 100 años (1912-2022): experiencias y posibilidades de educación matemática en América Latina. *Comuni@cción: Revista de Investigación en Comunicación y Desarrollo*, *14*(1). https://doi.org/10.33595/2226-1478.14.1.804

Wirawan, N., Syamsuri, Nindiasari, H., & Fatah, A. (2024). Meta-analysis: Ethnomathematics approaches in mathematical learning in Indonesia. *World Journal of Advanced Research and Reviews*, *22*(3), 832-839. https://doi.org/10.30574/wjarr.2024.22.3.1778

van Hiele, P. (1986). *Structure and insight: A theory of mathematics education.* Academic Press.

CAPÍTULO VIII

ESTRATEGIAS DIDÁCTICAS DE COMPROMISO INTERACTIVO Y CAMBIO CONCEPTUAL EN MECÁNICA CLÁSICA EN ESTUDIANTES DE SECUNDARIA DE CONTEXTOS PERIURBANOS DEL PERÚ, 2026

Judith Esther Gavidia Medrano

Universidad Nacional Hermilio Valdizán

Arnulfo Ortega Mallqui

Universidad Nacional Hermilio Valdizán

Resumen

Objetivo. Explicar por qué vías las estrategias didácticas de compromiso interactivo se asocian con la comprensión conceptual newtoniana de estudiantes de secundaria, mediante un modelo de ecuaciones estructurales que contrasta un mecanismo de dos eslabones —conflicto cognitivo productivo y regulación metacognitiva conceptual— y su equivalencia entre mujeres y varones. **Método.** Estudio cuantitativo explicativo de panel de dos ondas con 724 estudiantes de cuarto y quinto grado (362 mujeres y 362 varones) de 14 instituciones públicas de seis distritos periurbanos de Lima Metropolitana y el Callao, evaluados con una versión en español del Force Concept Inventory y escalas de cuatro ítems. Se estimaron modelos de medida y estructural, efectos indirectos con 2000 remuestreos bootstrap, invarianza de medida y análisis multigrupo por sexo. **Resultados.** El modelo de mediación serial parcial ajustó bien (CFI = 0,992; RMSEA = 0,021; SRMR = 0,026). El efecto indirecto total fue 0,129 (IC 95 % [0,087; 0,178]), 57,6 % del efecto total, y la vía serial 0,039 (IC 95 % [0,020; 0,062]). La comprensión aumentó 10,72 puntos porcentuales (d_z = 0,617). La medida fue invariante entre sexos, y el vínculo entre las estrategias y el conflicto cognitivo fue más intenso en las mujeres (β = 0,554 frente a 0,371; p = 0,008). **Conclusiones.** El aprovechamiento de las estrategias interactivas depende de que provoquen conflicto cognitivo y de que este se convierta en regulación metacognitiva; ese primer eslabón opera con más fuerza en las estudiantes, lo que sitúa el diseño didáctico como palanca de equidad en la enseñanza de la física.

Palabras clave: cambio conceptual; mecánica newtoniana; aprendizaje activo; conflicto cognitivo; metacognición; ecuaciones estructurales; equidad de género; educación secundaria.

Abstract

Objective. To explain through which pathways interactive-engagement teaching strategies are associated with secondary students' Newtonian conceptual

understanding, using a structural equation model that tests a two-link mechanism —productive cognitive conflict and conceptual metacognitive regulation— and its equivalence between girls and boys. Method. Explanatory quantitative two-wave panel study of 724 ninth- and tenth-grade students (362 girls and 362 boys) from 14 public schools in six peri-urban districts of Metropolitan Lima and Callao, assessed with a Spanish version of the Force Concept Inventory and four-item scales. Measurement and structural models, indirect effects with 2000 bootstrap resamples, measurement invariance and a multigroup analysis by sex were estimated. Results. The partial serial mediation model fit well (CFI = 0.992; RMSEA = 0.021; SRMR = 0.026). The total indirect effect was 0.129 (95% CI [0.087, 0.178]), 57.6% of the total effect, and the serial pathway was 0.039 (95% CI [0.020, 0.062]). Conceptual understanding rose 10.72 percentage points (d_z = 0.617). Measurement was invariant across sex, and the link between strategies and cognitive conflict was stronger among girls (β = 0.554 versus 0.371; p = .008). Conclusions. The conceptual benefit of interactive strategies depends on their provoking cognitive conflict and on that conflict turning into metacognitive regulation; this first link operates more strongly among female students, which positions instructional design as an equity lever in physics teaching.

Keywords: conceptual change; Newtonian mechanics; active learning; cognitive conflict; metacognition; structural equation modeling; gender equity; secondary education.

8.1 INTRODUCCIÓN

Las ideas intuitivas sobre fuerza y movimiento figuran entre las concepciones más resistentes de la educación científica. Los estudiantes llegan a la secundaria convencidos de que el movimiento requiere una fuerza que lo sostenga o de que el cuerpo más pesado ejerce la mayor fuerza en una interacción, y esas ideas persisten tras la instrucción formal (Aguirre y Aguirre, 2025; Hernández et al., 2023). La investigación ha acumulado evidencia de que las estrategias de cambio conceptual mejoran el aprendizaje científico con efectos grandes, aunque muy heterogéneos. La síntesis más amplia disponible, con 218 estudios primarios y 18 051

estudiantes, estima un efecto medio de g = 1,10, que se reduce a g = 0,93 tras el ajuste bayesiano por sesgo y cuyo intervalo de predicción abarca de 0,19 a 2,38 (Paçacı et al., 2024). Esa dispersión indica que saber que las estrategias funcionan no basta: falta explicar por qué funcionan en unas aulas y no en otras, y para quién.

El problema es especialmente agudo en los sistemas escolares donde la enseñanza de la física se apoya en la exposición y el ejercicio rutinario. En el Perú, los análisis multinivel de las evaluaciones internacionales muestran que las oportunidades de aprendizaje explican parte de las diferencias de desempeño científico entre estudiantes y escuelas (Bazán et al., 2022). Las zonas periurbanas de Lima Metropolitana y el Callao concentran estas tensiones: instituciones públicas de reciente expansión, aulas numerosas y una población estudiantil marcada por la migración interna y la diversidad lingüística.

La evidencia sobre la enseñanza activa es, con todo, abundante. Las síntesis de segundo orden sitúan el efecto de la enseñanza por indagación en d = 0,62 (Öztürk et al., 2022), las intervenciones de cambio conceptual producen efectos grandes sobre la comprensión de contenidos biológicos frente a la enseñanza tradicional (Aleknavičiūtė et al., 2023) y los estudios observacionales de aula documentan que la calidad de la indagación varía considerablemente entre docentes (Kersting et al., 2023). En la región, las revisiones de estrategias de aprendizaje activo en secundaria enmarcan estas prácticas como una vía de mejora (Aguirre y López, 2026). Sin embargo, tres vacíos limitan el alcance de esta evidencia.

1) El primero es de mecanismo. La mayor parte de los estudios compara condiciones de instrucción y mide la ganancia, pero no contrasta qué procesos cognitivos median entre la estrategia y el cambio conceptual. La teoría propone que el conflicto cognitivo —la experiencia de que una predicción propia choca con la evidencia— es la puerta de entrada al cambio, y que la reestructuración se consolida cuando el estudiante regula su propio razonamiento, proceso que la literatura sitúa en el plano de la metacognición y de la cognición epistémica en física (Fleur

et al., 2021; Ulu y Yerdelen, 2024). Rara vez se ha puesto a prueba ese encadenamiento como un modelo estructural con medida latente y alternativas comparadas.

2) El segundo vacío es de equidad. Se ha examinado poco si los mecanismos por los que la instrucción interactiva produce cambio conceptual operan con la misma intensidad en mujeres y varones, comparación que solo es válida si la medida es equivalente entre grupos. El tercero es de contexto: casi toda la evidencia procede de cursos universitarios de países de ingresos altos, mientras que la secundaria pública latinoamericana, donde las brechas socioeconómicas y territoriales son más marcadas, está poco representada.

La pregunta que guía el estudio es: ¿por qué vías se asocian las estrategias didácticas de compromiso interactivo con la comprensión conceptual newtoniana de estudiantes de secundaria, y operan esas vías del mismo modo en mujeres y varones? Los objetivos son:

a) Estimar la contribución de las estrategias a la comprensión conceptual al cierre de la unidad de mecánica, controlando la comprensión inicial;

b) Contrastar un mecanismo serial en el que el conflicto cognitivo productivo precede a la regulación metacognitiva conceptual; y

c) Examinar la invarianza de la medida y de las trayectorias estructurales entre sexos.

El aporte del estudio no es local. Los distritos periurbanos del Perú funcionan como un sitio de prueba exigente para una teoría general del cambio conceptual, en el que las condiciones de partida son desfavorables y la heterogeneidad de efectos que registra la literatura debería hacerse visible. Si el mecanismo se sostiene allí, gana generalidad; si opera con distinta intensidad por sexo, ofrece una explicación de proceso para las brechas de género en la física que trasciende el contexto de la muestra.

8.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

8.2.1 Compromiso interactivo y cambio conceptual

Se entiende por estrategias de compromiso interactivo el conjunto de prácticas docentes que obligan al estudiante a comprometerse con una predicción, a contrastarla con evidencia —demostraciones, simulaciones o datos— y a discutirla con sus pares antes de recibir la explicación formal (Hake, 1998; Kersting et al., 2023). La calidad de esas prácticas condiciona el compromiso del estudiante y sus ganancias de aprendizaje (Aguirre y López, 2026). Se espera, por tanto, una asociación positiva y directa entre las estrategias y la comprensión conceptual al cierre de la unidad.

H1. Las estrategias de compromiso interactivo (EIC) se asocian positivamente con la comprensión conceptual newtoniana (CCN) en el postest, controlando la comprensión inicial.

8.2.2 El conflicto cognitivo como puerta de entrada

El conflicto cognitivo es la estrategia de cambio conceptual más estudiada y la que concentra los efectos más grandes en la síntesis disponible (g = 1,10; k = 150; Paçacı et al., 2024). Desde el punto de vista del estudiante, el conflicto es una experiencia —la sorpresa ante un resultado inesperado y la necesidad de revisar la idea propia— que las prácticas interactivas deberían provocar con más frecuencia que la exposición.

H2. Las EIC se asocian positivamente con el conflicto cognitivo productivo (CCG).

8.2.3 La regulación metacognitiva como consolidación

La experiencia del conflicto no garantiza el cambio: puede resolverse descartando la evidencia. La reestructuración requiere que el estudiante supervise la coherencia de su explicación, identifique qué parte de su razonamiento debe cambiar y la relacione con situaciones nuevas, procesos propios de la metacognición (Fleur et al., 2021) que, en física, se vinculan

con la cognición epistémica (Ulu y Yerdelen, 2024). Se propone que el conflicto activa esa regulación y que ambos contribuyen a la comprensión conceptual.

H3. El CCG se asocia positivamente con la regulación metacognitiva conceptual (MET).

H4. La MET (H4a) y el CCG (H4b) se asocian positivamente con la CCN, controlando la comprensión inicial.

H5. Las EIC se asocian con la CCN de forma indirecta a través del CCG (H5a), de la MET (H5b) y de la secuencia CCG → MET (H5c).

8.2.4 Diferencias por sexo en el mecanismo

Si las mujeres parten de una comprensión inicial más baja, cabe esperar que el beneficio de las estrategias se transmita con más fuerza en ellas a través del primer eslabón del mecanismo: la discusión entre pares y la confrontación con la evidencia ofrecen un espacio de baja amenaza para reconocer que una idea propia no funciona. Esa comparación solo es válida si la medida es equivalente entre grupos.

H6. La medida de los cuatro constructos es invariante entre mujeres y varones (H6a), y la asociación EIC → CCG es más intensa en las mujeres (H6b).

8.3 MATERIALES Y MÉTODOS

8.3.1 Diseño y participantes

Estudio cuantitativo, explicativo y de panel de dos ondas. La comprensión conceptual se midió al inicio (marzo y abril de 2026) y al cierre (julio de 2026) de la unidad de mecánica; las percepciones sobre las estrategias, el conflicto cognitivo y la regulación metacognitiva se midieron en la segunda onda, referidas a la unidad completa. La comprensión inicial entró como covariable en todos los modelos. El diseño no permite atribuir causalidad: las relaciones estimadas son asociaciones temporales y estructurales.

Participaron 724 estudiantes de cuarto y quinto grado de 14 instituciones públicas y 28 aulas de seis distritos periurbanos de Lima Metropolitana y el Callao: Carabayllo (n = 156), Puente Piedra (n = 104), Ventanilla (n = 104), Ate (n = 153), Villa María del Triunfo (n = 104) y Pachacámac (n = 103). La edad media fue de 15,55 años (DE = 0,77). El muestreo fue no probabilístico por conveniencia, con cuotas por sexo en cada aula, de modo que los resultados describen a los estudiantes analizados y no a la población escolar del país.

El tamaño muestral se fijó a priori para el análisis más exigente, la comparación multigrupo de trayectorias entre sexos. Con potencia de 0,80, α bilateral de 0,05 y correlaciones de 0,30 y 0,10 ancladas en la literatura ($q = 0{,}2092$), $n_g = 2[(1{,}960 + 0{,}842)/q]^2 + 3 = 361{,}74$, redondeado a 362 por grupo. Ese N cubre el contraste de ajuste aproximado con 95 grados de libertad ($\varepsilon_0 = 0{,}05$ frente a $\varepsilon_1 = 0{,}08$), que exige $N \geq 137$ (MacCallum et al., 1996; Moshagen y Bader, 2024). Una simulación de Monte Carlo con 300 réplicas de N = 724, sobre trayectorias en el extremo conservador de la literatura, estimó una potencia de 0,98 para la vía serial, 0,85 para la vía metacognitiva y 0,55 para la vía a través del conflicto cognitivo (Wang y Rhemtulla, 2021).

8.3.2 Instrumentos y procedimiento

La comprensión conceptual se midió con una versión en español del Force Concept Inventory (Hestenes et al., 1992), de 30 ítems, aplicada en las dos ondas y agrupada por contenido en tres indicadores de porcentaje de aciertos: cinemática (ocho ítems), dinámica (catorce) e interacción (ocho), en la línea de los análisis por contenido de los inventarios de mecánica (Le et al., 2025). Las estrategias, el conflicto cognitivo y la regulación metacognitiva se midieron con cuatro ítems cada uno en escala de 1 a 5, según muestra la Tabla 8-1.

Tabla 8-1

Operacionalización de los constructos

Constructo	Código	Indicador	Escala
Estrategias de compromiso interactivo (EIC)	e1	El docente nos pide predecir el resultado de un fenómeno antes de observarlo	Likert 1–5
	e2	Discutimos en parejas nuestras respuestas a preguntas conceptuales antes de la explicación del docente	
	e3	Contrastamos nuestras predicciones con demostraciones o simulaciones en clase	
	e4	El docente usa nuestras respuestas para ajustar la explicación de la clase	
Conflicto cognitivo productivo (CCG)	c1	Me sorprendió que el resultado de una demostración contradijera lo que yo esperaba	
	c2	Sentí la necesidad de revisar mi idea inicial cuando no coincidía con la evidencia	
	c3	Las discusiones con mis compañeros me hicieron dudar de mis explicaciones previas	
	c4	Busqué una explicación nueva cuando mi idea sobre el movimiento no funcionaba	
Regulación metacognitiva conceptual (MET)	m1	Compruebo si mi explicación de un fenómeno es coherente con las leyes de Newton	
	m2	Identifico qué parte de mi razonamiento debo cambiar después de resolver un problema	
	m3	Reviso mis respuestas conceptuales antes de darlas por correctas	
	m4	Relaciono lo aprendido en clase con situaciones de movimiento de mi vida diaria	
Comprensión conceptual newtoniana (CCN)	k_cin	Cinemática: posición, velocidad y aceleración (8 ítems)	% de aciertos, postest
	k_din	Dinámica: primera y segunda leyes, fuerza neta (14 ítems)	
	k_int	Interacción: tercera ley y tipos de fuerza (8 ítems)	
Comprensión conceptual inicial (covariable)	pre	Puntuación total del inventario en el pretest (30 ítems)	

Nota: Ítems de percepción referidos a la unidad de mecánica y aplicados en la segunda onda. Indicadores de la CCN obtenidos al agrupar por contenido los 30 ítems del *Force Concept Inventory* en español..

La aplicación se realizó en el aula, en horario regular. El inventario se aplicó en la primera y la última semana de la unidad, con 45 minutos para responder. Para atenuar la varianza de método común se separaron las escalas y el inventario, se garantizó el anonimato y se evaluó el sesgo con el factor único de Harman y los factores de inflación de la varianza (Podsakoff et al., 2024).

8.3.3 Estrategia analítica

Se estimaron un modelo de medida de cuatro factores y el modelo estructural hipotetizado por máxima verosimilitud. Se evaluaron la fiabilidad (α y ω), la validez convergente (cargas y AVE) y la discriminante (HTMT; Henseler et al., 2015, y raíz de la AVE). El ajuste se valoró con los criterios de Hu y Bentler (1999): CFI y TLI ≥ 0,95, RMSEA ≤ 0,06 y SRMR ≤ 0,08. El modelo se comparó con tres alternativas: sin mediación (M0), mediación completa (M2) y mediadores paralelos sin la vía CCG → MET (M3), mediante $\Delta\chi^2$ y criterios de información (Li y Singh, 2023). Los intervalos se obtuvieron con 2000 remuestreos bootstrap percentil, y los efectos indirectos se repitieron con 500 remuestreos por aulas.

La invarianza entre sexos se contrastó en tres niveles —configural, métrico y escalar— con $\Delta\chi^2$, ΔCFI ≥ −0,010 y ΔRMSEA ≤ 0,015 (Leitgöb et al., 2023); omitir este contraste compromete la comparación entre grupos (Maassen et al., 2025; Navarro et al., 2024), y la invarianza del inventario cuenta con evidencia previa (Buncher et al., 2025; Morley et al., 2023). Con invarianza métrica, cada trayectoria se comparó fijándola igual entre grupos ($\Delta\chi^2$ con un grado de libertad). El cambio conceptual se describió con la prueba t relacionada, la prueba de Wilcoxon, d_z y la ganancia normalizada de Hake (Christman et al., 2024; Coletta, 2023; Hake, 1998). Los análisis se ejecutaron en Python 3.12 con semopy 2.3.11 y SciPy 1.15.3; el estimador multigrupo se programó para este estudio y reproduce la suma de los ajustes por grupo (χ^2 = 236,43 frente a 236,44).

8.4 RESULTADOS

8.4.1 Características de la muestra

La Tabla 8-2 resume las características por sexo. Mujeres y varones no difirieron en edad, grado ni escolaridad materna; la proporción con lengua materna originaria fue mayor entre las mujeres, que además obtuvieron una comprensión conceptual inicial y final más baja.

Tabla 8-2

Características de los participantes por sexo

Variable		Mujeres (n = 362)	Varones (n = 362)	Contraste	p
Edad en años, M (DE)		15,55 (0,78)	15,56 (0,76)	t = -0,11	0,912
Quinto grado, n (%)		181 (50,0)	181 (50,0)	$\chi^2(1) = 0,00$	1,000
Lengua materna originaria, n (%)		78 (21,5)	53 (14,6)	$\chi^2(1) = 5,82$	0,016
Madre con secundaria completa, n (%)		216 (59,7)	218 (60,2)	$\chi^2(1) = 0,02$	0,879
Comprensión conceptual, % de aciertos, M (DE)	inicial	28,22 (19,46)	32,01 (20,03)	t = -2,58; d = -0,192	0,010
	final	38,47 (18,97)	43,20 (19,37)	t = -3,32; d = -0,247	< 0,001

Nota: t = prueba t de Welch; χ^2 = prueba de independencia de Pearson; d = diferencia estandarizada (mujeres menos varones). Comprensión conceptual medida con la versión en español del Force Concept Inventory (30 ítems).

8.4.2 Cambio conceptual entre el pretest y el postest

La comprensión conceptual total aumentó 10,72 puntos porcentuales (IC 95 % [9,45; 11,99]; t(723) = 16,60; p < 0,001; d_z = 0,617), con una ganancia normalizada media de 0,153. El aumento fue significativo en las tres dimensiones y en ambos sexos (Tabla 8-3; Figura 8-1).

Tabla 8-3

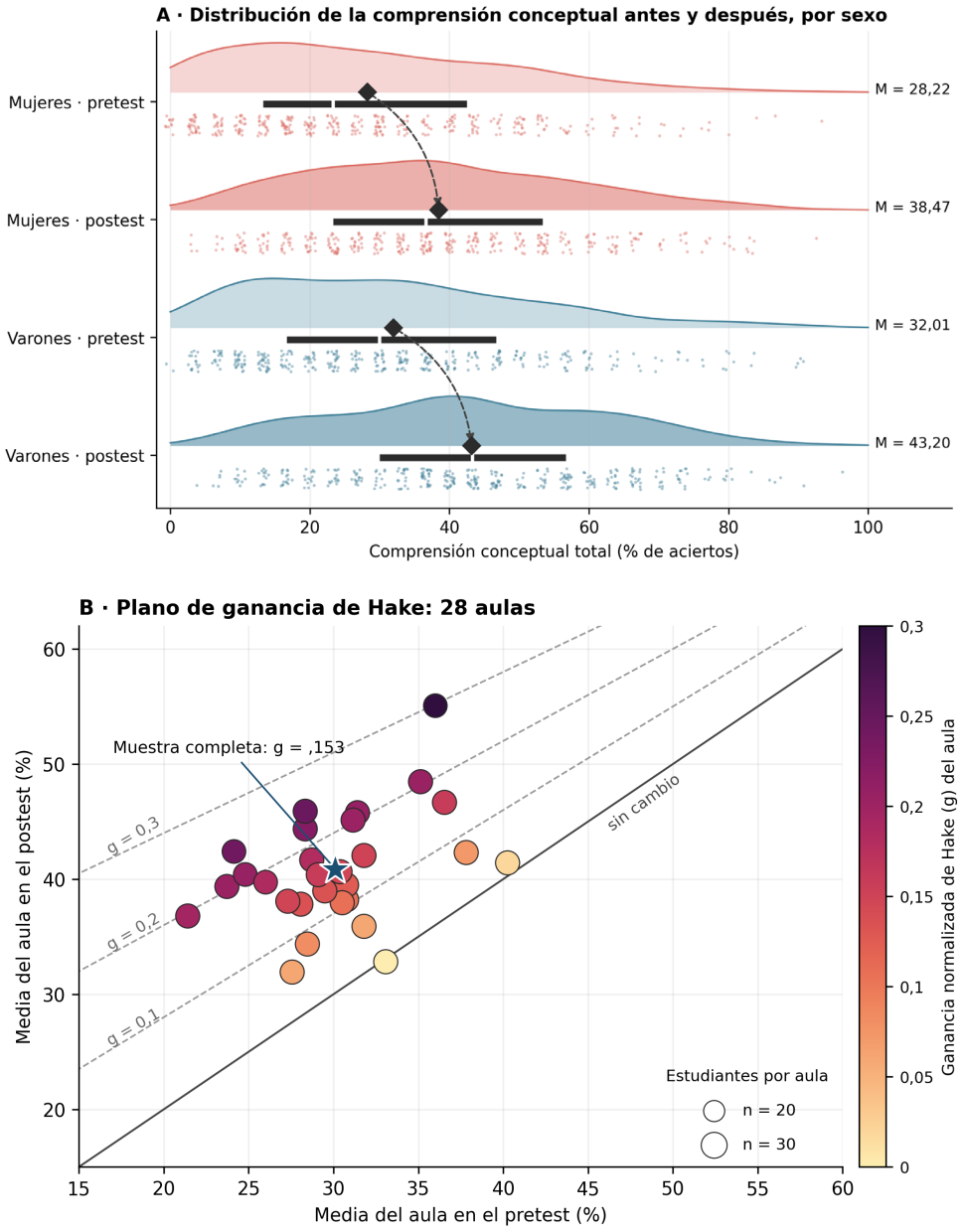
Comprensión conceptual al inicio y al cierre de la unidad de mecánica (porcentaje de aciertos)

Dimensión	n	Pretest M (DE)	Postest M (DE)	Diferencia [IC 95 %]	t (gl)	p	d_z	<g>
Cinemática	724	30,21 (23,47)	42,16 (24,30)	11,95 [9,99; 13,91]	11,95 (723)	< 0,001	0,444	0,171
Dinámica (primera y segunda leyes)	724	26,59 (22,48)	36,22 (23,16)	9,63 [7,94; 11,31]	11,22 (723)	< 0,001	0,417	0,131
Interacción (tercera ley)	724	36,19 (26,14)	47,60 (26,45)	11,41 [9,24; 13,58]	10,33 (723)	< 0,001	0,384	0,179
Total	724	30,12 (19,82)	40,84 (19,30)	10,72 [9,45; 11,99]	16,60 (723)	< 0,001	0,617	0,153
Total — mujeres	362	28,22 (19,46)	38,47 (18,97)	10,25 [8,47; 12,03]	11,32 (361)	< 0,001	0,595	0,143
Total — varones	362	32,01 (20,03)	43,20 (19,37)	11,20 [9,38; 13,01]	12,14 (361)	< 0,001	0,638	0,165

Nota: t = prueba t para muestras relacionadas; la prueba de rangos con signo de Wilcoxon dio p < 0,001 en todas las filas. d_z = diferencia media dividida por la desviación típica de las diferencias. <g> = ganancia normalizada de Hake, (postest – pretest)/(100 – pretest).

Figura 8-1

Cambio conceptual entre el pretest y el postest en mecánica newtoniana



Nota. N = 724; puntuaciones en porcentaje de aciertos. Panel A: densidad, observaciones individuales, mediana e intervalo intercuartílico de la puntuación total por sexo y momento; los rombos marcan las medias (mujeres, de 28,22 a 38,47; varones, de 32,01 a 43,20). Panel B: media del pretest frente a media del postest en cada una de las 28 aulas, con isolíneas de ganancia normalizada de Hake (g = 0,1, 0,2 y 0,3) y la diagonal de ausencia de cambio; el tamaño indica los estudiantes por aula y el color la g del aula, que osciló entre –0,004 y 0,298 (mediana 0,155) y fue positiva en 27 de las 28 aulas; la estrella marca la muestra completa (g = ,153).

Nota. N = 724; puntuaciones en porcentaje de aciertos. Panel C: densidad del cambio individual (postest menos pretest) por dimensión, con la diferencia media y su IC 95 %: cinemática 11,95 [9,99; 13,91], d_z = ,444, g = ,171; dinámica 9,63 [7,94; 11,31], d_z = ,417, g = ,131; interacción 11,41 [9,24; 13,58], d_z = ,384, g = ,179; total 10,72 [9,45; 11,99], d_z = ,617, g = ,153; mejoraron el 58,8 %, el 59,0 %, el 56,8 % y el 72,2 % de los estudiantes, respectivamente.

8.4.3 Modelo de medida

El modelo de medida ajustó a los datos ($\chi^2(95) = 126,65$; CFI = 0,992; TLI = 0,990; RMSEA = 0,021, IC 90 % [0,010; 0,031]; SRMR = 0,026). Todas las cargas estandarizadas fueron significativas ($p < 0,001$) y superiores a 0,59. Los tres constructos de percepción alcanzaron α y ω superiores a 0,80 y AVE superiores a 0,50 (Tabla 8-4); la comprensión conceptual obtuvo una fiabilidad compuesta de 0,689 y una AVE de 0,427.

Tabla 8-4

Estadísticos descriptivos, cargas factoriales, fiabilidad y validez convergente

Constructo	Indicador	M	DE	λ	α	ω	AVE
Estrategias de compromiso interactivo (EIC)	e1	3,13	1,04	0,751	0,841	0,842	0,571
	e2	3,12	1,05	0,773			
	e3	3,08	1,06	0,788			
	e4	3,11	1,04	0,708			
Conflicto cognitivo productivo (CCG)	c1	3,07	1,06	0,691	0,815	0,816	0,527
	c2	3,11	1,00	0,767			
	c3	3,11	1,06	0,688			
	c4	3,12	1,02	0,752			
Regulación metacognitiva conceptual (MET)	m1	3,10	1,04	0,747	0,811	0,812	0,520
	m2	3,13	1,03	0,712			
	m3	3,09	1,03	0,760			
	m4	3,15	1,05	0,662			
Comprensión conceptual newtoniana (CCN)	k_cin	42,16	24,30	0,614	0,685	0,689	0,427
	k_din	36,22	23,16	0,740			
	k_int	47,60	26,45	0,596			

Nota: λ = carga estandarizada; α = alfa de Cronbach; ω = omega de McDonald (fiabilidad compuesta); AVE = varianza media extraída. Indicadores de la CCN en porcentaje de aciertos del postest. Todas las cargas, $p < 0,001$.

La asimetría y la curtosis de los indicadores se situaron entre -0,99 y 0,45, y la curtosis multivariada fue ligeramente inferior a la esperada bajo normalidad ($b_{2,p}$ = 283,38 frente a 288; z = −2,59).

La validez discriminante se sostuvo por los dos criterios: el HTMT máximo fue 0,499, entre el conflicto cognitivo y la regulación metacognitiva, y la raíz de la AVE de cada constructo superó sus correlaciones con los demás (Tabla 8-5). El primer factor de Harman explicó el 33,2 % de la varianza de los indicadores de percepción y los factores de inflación de la varianza se situaron entre 1,213 y 1,356.

Tabla 8-5

Validez discriminante: correlaciones entre factores, raíz de la AVE y razón HTMT

Constructo	EIC	CCG	MET	CCN
EIC	**0,756**	(0,477)	(0,352)	(0,366)
CCG	0,478	**0,726**	(0,499)	(0,412)
MET	0,350	0,502	**0,721**	(0,475)
CCN	0,377	0,410	0,472	**0,653**

Nota: Diagonal (negrita): raíz cuadrada de la AVE. Bajo la diagonal: correlaciones entre factores del modelo de medida. Sobre la diagonal, entre paréntesis: razón heterorrasgo-monorrasgo (HTMT). La raíz de la AVE de cada constructo supera todas sus correlaciones con los demás, y el HTMT máximo es 0,499.

8.4.4 Modelo estructural y comparación con modelos alternativos

El modelo hipotetizado (M1) ajustó igual que el de medida, porque especifica todas las trayectorias entre factores. Eliminar la mediación (M0) empeoró el ajuste de forma marcada ($\Delta\chi^2(4) = 177{,}30$; $p < 0{,}001$; SRMR = 0,116), eliminar el efecto directo (M2) también ($\Delta\chi^2(1) = 4{,}54$; $p = 0{,}033$) y eliminar la vía CCG → MET (M3) produjo el mayor deterioro entre las alternativas con un grado de libertad ($\Delta\chi^2(1) = 70{,}16$; $p < 0{,}001$). M1 obtuvo el menor AIC (Tabla 8-6; Figura 8-2).

Tabla 8-6

Índices de ajuste del modelo de medida, del modelo hipotetizado y de los modelos alternativos

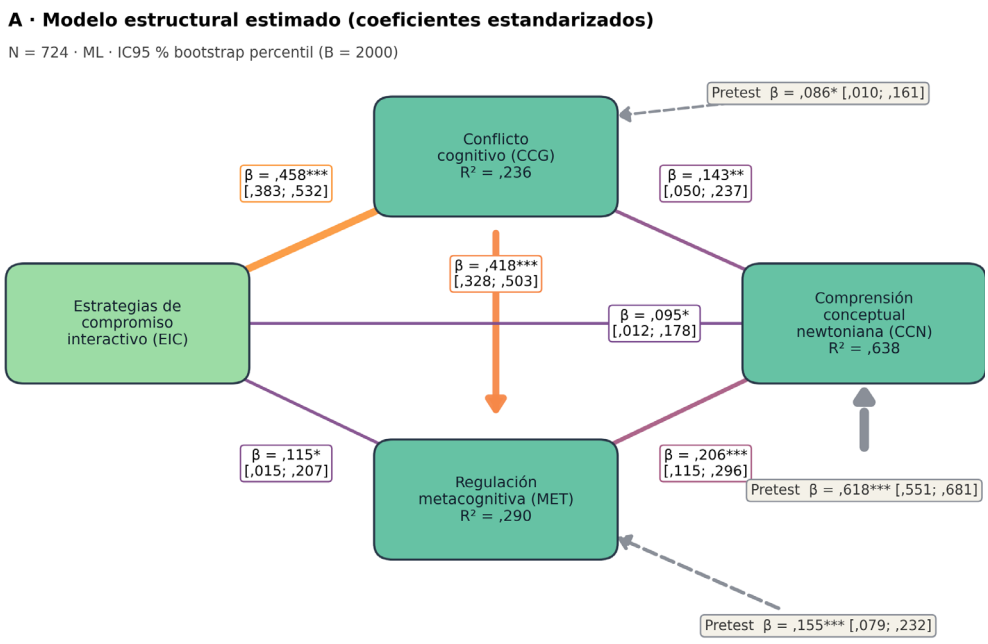
Modelo		χ^2	gl	χ^2/gl	CFI	TLI	RMSEA [IC 90 %]	SRMR	AIC	BIC	$\Delta\chi^2$ (Δgl) frente a M1	p
Medida:	cuatro factores y covariable	126,65	95	1,33	0,992	0,990	0,021 [0,010; 0,031]	0,026	34434,4	34622,4	—	—
M1:	mediación serial parcial (hipotetizado)	126,65	95	1,33	0,992	0,990	0,021 [0,010; 0,031]	0,026	34434,4	34622,4	—	—
M0:	sin mediación	303,95	99	3,07	0,950	0,939	0,054 [0,047; 0,060]	0,116	34603,9	34773,6	177,30 (4)	< 0,001
M2:	mediación completa (sin EIC → CCN)	131,19	96	1,37	0,991	0,989	0,023 [0,011; 0,032]	0,027	34436,9	34620,3	4,54 (1)	0,033
M3:	mediadores paralelos (sin CCG → MET)	196,81	96	2,05	0,975	0,969	0,038 [0,030; 0,046]	0,060	34502,7	34686,0	70,16 (1)	< 0,001

Nota: Estimación por máxima verosimilitud; N = 724. Criterios de Hu y Bentler (1999): CFI y TLI ≥ 0,95; RMSEA ≤ 0,06; SRMR ≤ 0,08. El modelo de medida y M1 tienen el mismo ajuste porque M1 especifica todas las trayectorias entre factores.

Las estrategias se asociaron con el conflicto cognitivo (β = 0,458), y este con la regulación metacognitiva (β = 0,418), lo que apoya H2 y H3. Controlando la comprensión inicial (β = 0,618), la regulación metacognitiva (β = 0,206) y el conflicto cognitivo (β = 0,143) se asociaron con la comprensión final, lo que apoya H4a y H4b. El efecto directo de las estrategias fue β = 0,095, lo que apoya H1 con un efecto pequeño. El modelo explicó el 23,6 % de la varianza del conflicto cognitivo, el 29,0 % de la regulación metacognitiva y el 63,8 % de la comprensión final (Tabla 8-7). Sin contar la comprensión inicial, la contribución única de la regulación metacognitiva (f^2 = 0,064) superó la del conflicto cognitivo (f^2 = 0,021) y la de las estrategias (f^2 = 0,009).

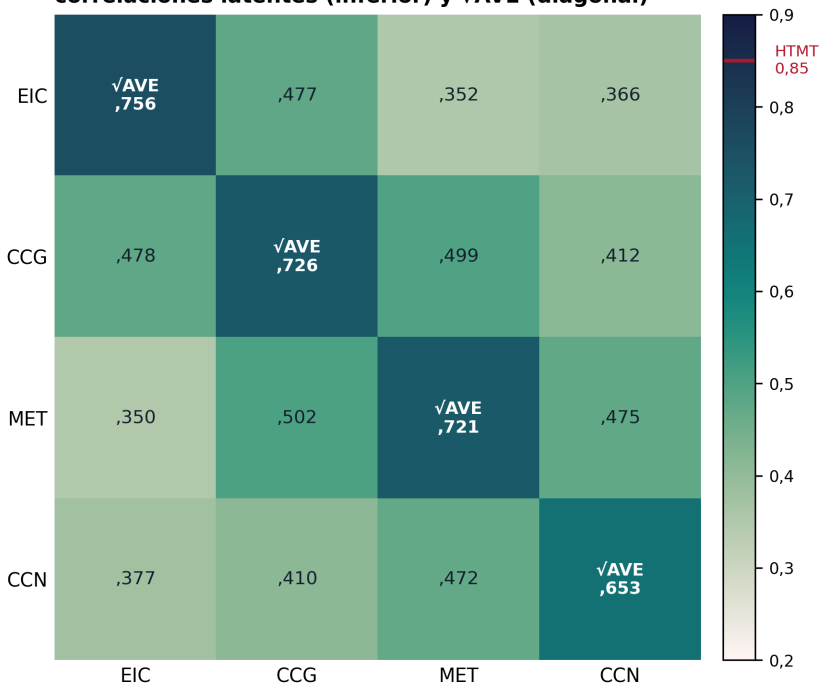
Figura 8-2

Modelo de medida y modelo estructural de la mediación serial del cambio conceptual



Nota. N = 724 estudiantes de 28 aulas; estimación por máxima verosimilitud. Panel A: coeficientes estandarizados (β) con su IC 95 % bootstrap percentil (B = 2000); el grosor de cada trayectoria es proporcional a β y la comprensión inicial (pretest) entra como covariable de los tres constructos endógenos; R^2 = ,236 para el conflicto cognitivo, ,290 para la regulación metacognitiva y ,638 para la comprensión conceptual newtoniana. * p < ,05; ** p < ,01; *** p < ,001.

B · Validez discriminante: HTMT (triángulo superior), correlaciones latentes (inferior) y √AVE (diagonal)

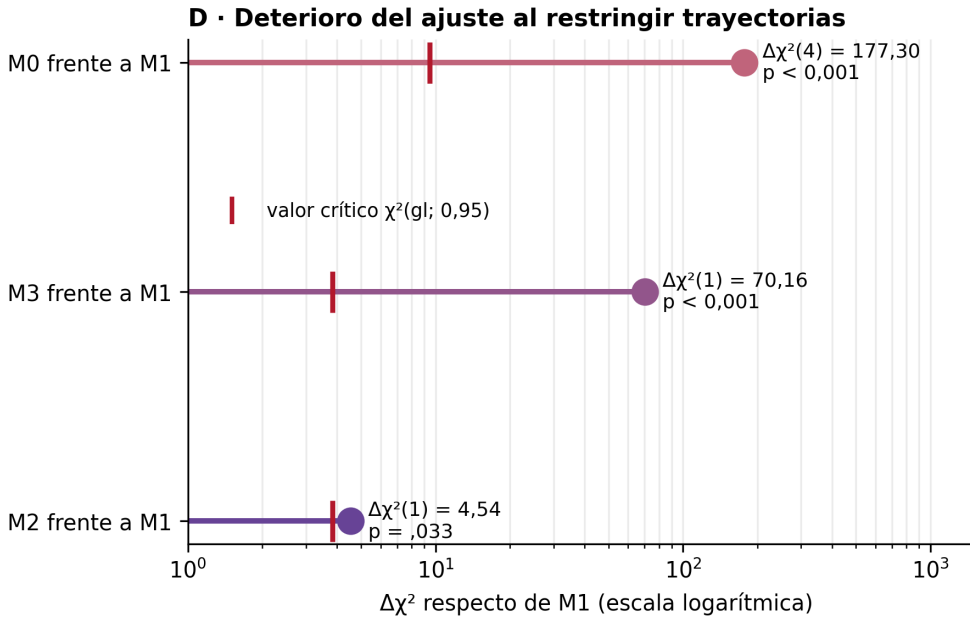


C · Índices de ajuste del modelo hipotetizado y de los modelos rivales

	χ^2/gl	CFI	TLI	RMSEA	SRMR
M1 · hipotetizado	**1,33**	**,992**	**,990**	**,021**	**,026**
M2 · mediación completa	1,37	,991	,989	,023	,027
M3 · mediadores paralelos	2,05	,975	,969	,038	,060
M0 · solo efectos directos	3,07	,950	,939	,054	,116

Celda verde: dentro del criterio (χ^2/gl ≤ 3; CFI y TLI ≥ 0,95; RMSEA ≤ 0,06; SRMR ≤ 0,08; ΔAIC ≤ 2 respecto de M1).

Nota. N = 724 estudiantes de 28 aulas; estimación por máxima verosimilitud. Panel B: validez discriminante: HTMT en el triángulo superior (máximo ,499), correlaciones latentes en el inferior y raíz cuadrada de la AVE en la diagonal; la marca roja de la escala señala el criterio HTMT de 0,85. Panel C: índices de ajuste del modelo hipotetizado (M1: CFI = ,992; TLI = ,990; RMSEA = ,021; SRMR = ,026) y de los tres modelos rivales; en verde, los valores dentro del criterio; ΔAIC respecto de M1.



Nota. N = 7 Nota. N = 724 estudiantes de 28 aulas; estimación por máxima verosimilitud. Panel D: deterioro del ajuste al restringir trayectorias, en escala logarítmica: M0 sin mediación, $\Delta\chi^2(4) = 177{,}30$; M3 con mediadores paralelos, $\Delta\chi^2(1) = 70{,}16$; M2 sin efecto directo, $\Delta\chi^2(1) = 4{,}54$, p = ,033; la marca roja señala el valor crítico de χ^2 con $\alpha = 0{,}05$.

8.4.5 Efectos indirectos

Las tres vías indirectas fueron distintas de cero, lo que apoya H5a, H5b y H5c (Tabla 8-8; Figura 8-3). El efecto indirecto total representó el 57,6 % del efecto total de las estrategias sobre la comprensión final (0,224; IC 95 % [0,150; 0,298]). Con remuestreo por aulas, los intervalos del efecto indirecto total ([0,082; 0,179]) y de la vía serial ([0,025; 0,056]) se mantuvieron alejados del cero; la percepción de las estrategias mostró una correlación intraclase de 0,119 (efecto de diseño = 3,96).

Tabla 8-7

Coeficientes del modelo estructural hipotetizado

Trayectoria	Hipótesis	B	EE	z	p	β	IC 95 % de β (bootstrap)	Decisión
EIC → CCG	H2	0,430	0,046	9,44	< 0,001	0,458	[0,383; 0,532]	Apoyada
CCG → MET	H3	0,443	0,057	7,81	< 0,001	0,418	[0,328; 0,503]	Apoyada
EIC → MET	—	0,114	0,048	2,36	0,018	0,115	[0,015; 0,207]	Significativa
MET → CCN	H4a	0,396	0,092	4,30	< 0,001	0,206	[0,115; 0,296]	Apoyada
CCG → CCN	H4b	0,292	0,103	2,85	0,004	0,143	[0,050; 0,237]	Apoyada
EIC → CCN	H1	0,181	0,085	2,14	0,033	0,095	[0,012; 0,178]	Apoyada
Pretest → CCG	Control	0,032	0,014	2,21	0,027	0,086	[0,010; 0,161]	Significativa
Pretest → MET	Control	0,061	0,015	4,04	< 0,001	0,155	[0,079; 0,232]	Significativa
Pretest → CCN	Control	0,466	0,035	13,38	< 0,001	0,618	[0,551; 0,681]	Significativa
R^2						CCG = 0,236 · MET = 0,290 · CCN = 0,638		

Nota: B = coeficiente no estandarizado; EE = error estándar de máxima verosimilitud; β = coeficiente estandarizado; IC 95 % percentil con 2000 remuestreos bootstrap. Pretest = comprensión conceptual inicial (covariable). La decisión se basa en si el intervalo bootstrap excluye el cero.

Tabla 8-8

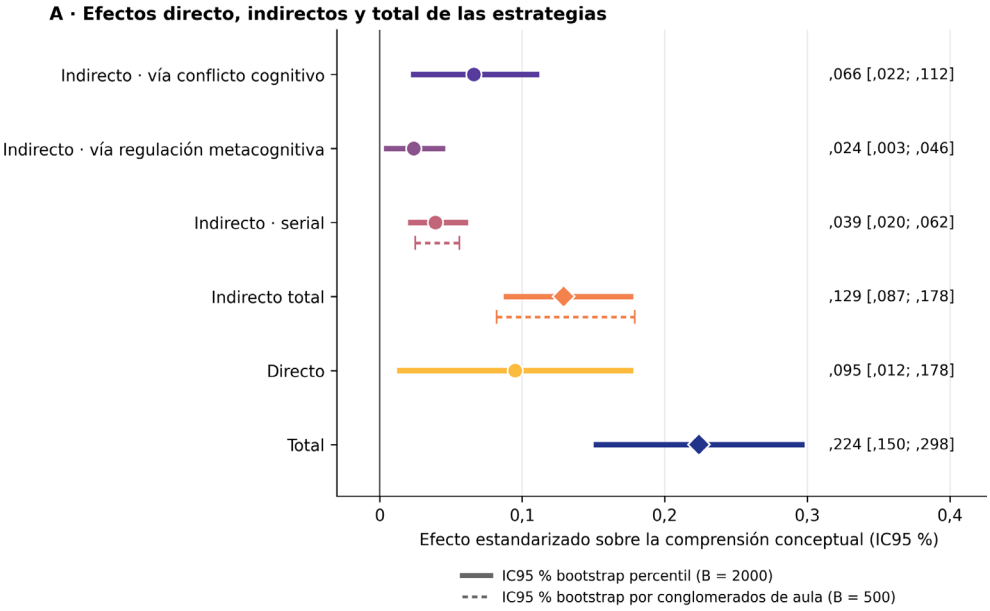
Descomposición del efecto de las estrategias de compromiso interactivo sobre la comprensión conceptual final

Efecto	Hipótesis	Estimación estandarizada	EE bootstrap	IC 95 % percentil	Decisión
Indirecto: EIC → CCG → CCN	H5a	0,066	0,023	[0,022; 0,112]	Distinto de cero
Indirecto: EIC → MET → CCN	H5b	0,024	0,011	[0,003; 0,046]	
Indirecto serial: EIC → CCG → MET → CCN	H5c	0,039	0,011	[0,020; 0,062]	
Indirecto total	—	0,129	0,023	[0,087; 0,178]	
Directo: EIC → CCN	H1	0,095	0,043	[0,012; 0,178]	
Total	—	0,224	0,039	[0,150; 0,298]	
Proporción mediada		0,576			

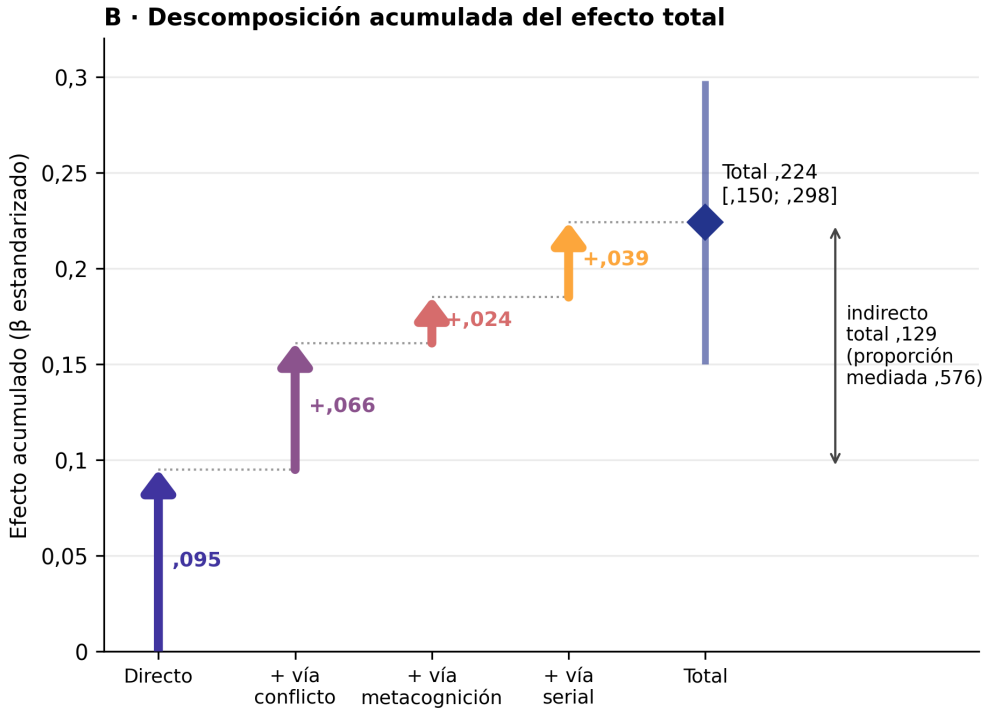
Nota: Efectos estandarizados; 2000 remuestreos bootstrap percentil. Proporción mediada = efecto indirecto total / efecto total.

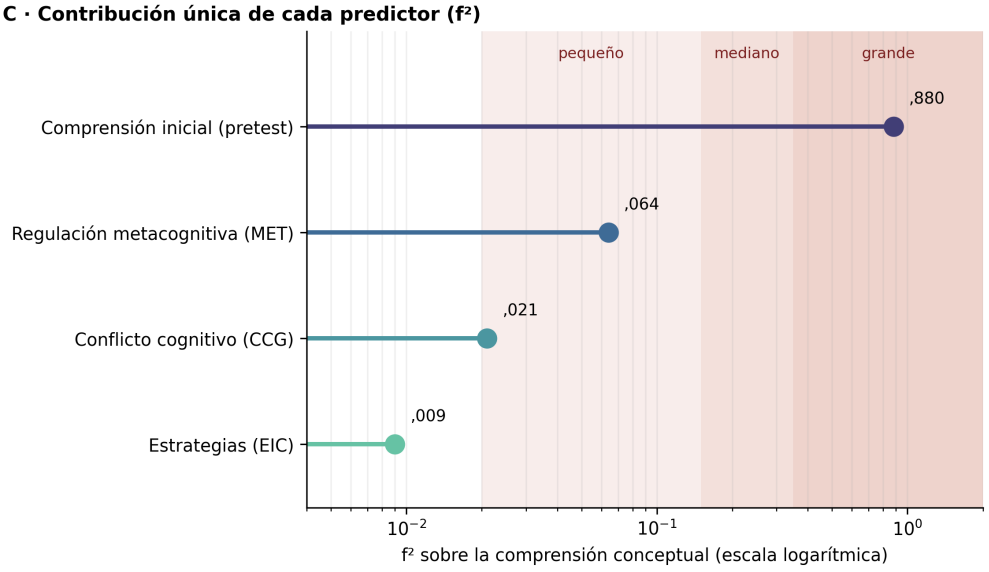
Figura 8-3

Descomposición del efecto de las estrategias de compromiso interactivo sobre la comprensión conceptual newtoniana

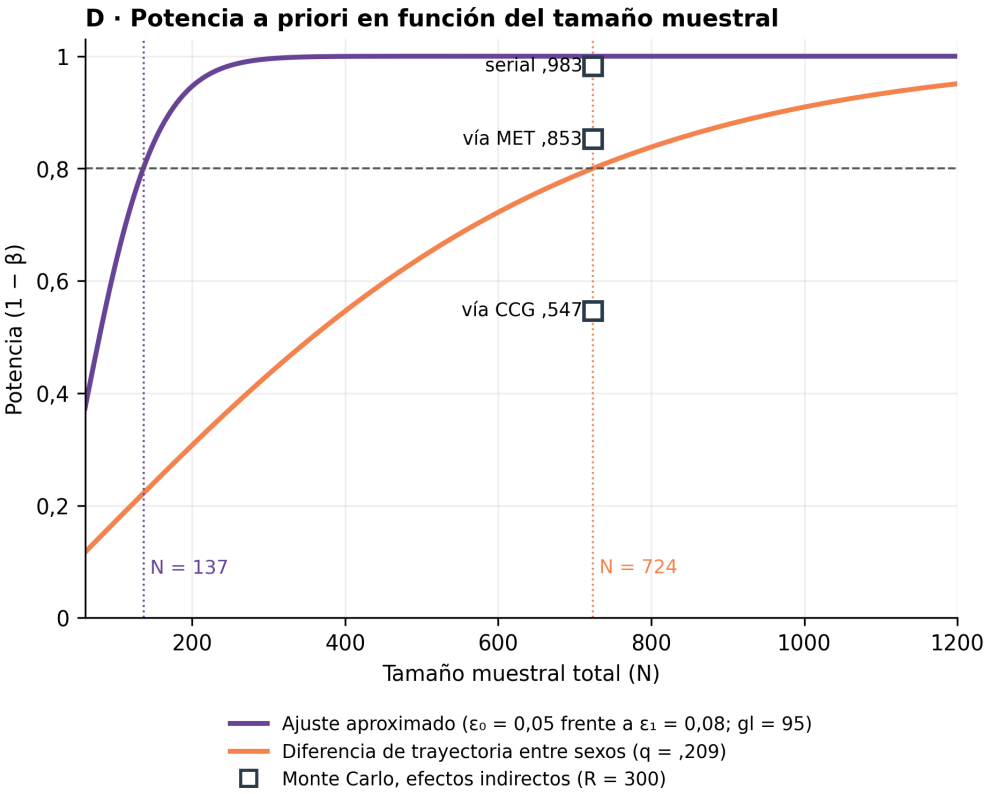


Nota. N = 724; efectos estandarizados. Panel A: efectos con IC 95 % bootstrap percentil (B = 2000; línea continua) y, para la vía serial y el indirecto total, IC 95 % bootstrap por conglomerados de aula (B = 500; línea discontinua): vía conflicto cognitivo ,066 [,022; ,112]; vía regulación metacognitiva ,024 [,003; ,046]; vía serial ,039 [,020; ,062], por conglomerados [,025; ,056]; indirecto total ,129 [,087; ,178], por conglomerados [,082; ,179]; directo ,095 [,012; ,178]; total ,224 [,150; ,298].





Nota. N = 724; efectos estandarizados. Panel B: suma acumulada del efecto directo y de las tres vías indirectas hasta el efecto total; proporción mediada = ,576. Panel C: contribución única (f^2) de cada predictor de la comprensión conceptual final, en escala logarítmica, sobre las bandas convencionales de efecto pequeño, mediano y grande: pretest ,88; regulación metacognitiva ,064; conflicto cognitivo ,021; estrategias ,009.



Nota. N = 724; efectos estandarizados. Panel D: potencia a priori en función de N para el contraste de ajuste aproximado (gl = 95; $\varepsilon_0 = 0{,}05$ frente a $\varepsilon_1 = 0{,}08$; $\alpha = 0{,}05$), que alcanza 0,80 con N = 137, y para el contraste de una trayectoria entre sexos (q = 0,2092), que la alcanza con 362 estudiantes por grupo (N = 724); los cuadrados marcan la potencia de Monte Carlo de las vías indirectas con N = 724 (R = 300): vía serial ,983; vía metacognitiva ,853; vía conflicto ,547.

8.4.6 Invarianza de medida y análisis multigrupo por sexo

El modelo configural ajustó en ambos grupos (CFI = 0,989; RMSEA = 0,026; SRMR = 0,034). Igualar las cargas y luego los interceptos no deterioró el ajuste, por lo que la medida fue invariante en los niveles métrico y escalar (H6a; Tabla 8-9).

Tabla 8-9

Invarianza de medida entre mujeres y varones

Nivel	χ^2	gl	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ (Δgl)	p
Configural	236,43	190	0,989	0,986	0,026	0,034	—	—
Métrica (cargas iguales)	245,34	201	0,989	0,987	0,025	0,038	8,91 (11)	0,630
Escalar (cargas e interceptos iguales)	260,86	212	0,988	0,987	0,025	0,038	15,52 (11)	0,160

Nivel	ΔCFI	ΔRMSEA	Decisión
Configural	—	—	Referencia
Métrica (cargas iguales)	0,000	-0,001	Se sostiene
Escalar (cargas e interceptos iguales)	-0,001	0,000	Se sostiene

Nota: Máxima verosimilitud multigrupo; mujeres n = 362, varones n = 362. Criterios: ΔCFI ≥ –0,010 y ΔRMSEA ≤ 0,015.

Con invarianza métrica, 2 de las nueve trayectorias difirieron entre sexos: EIC → CCG (β = 0,554 en mujeres frente a 0,371 en varones; $\Delta\chi^2(1) = 6,95$; p = 0,008), lo que apoya H6b, y Pretest → CCG (β = 0,198 frente a -0,025; $\Delta\chi^2(1) = 8,84$; p = 0,003) (Tabla 8-10; Figura 8-4). La vía indirecta a través del conflicto cognitivo fue mayor en las mujeres (0,122 frente a 0,040), al igual que la vía serial (0,064 frente a 0,023). Las demás trayectorias no difirieron significativamente.

Tabla 8-10

Análisis multigrupo de las trayectorias estructurales por sexo

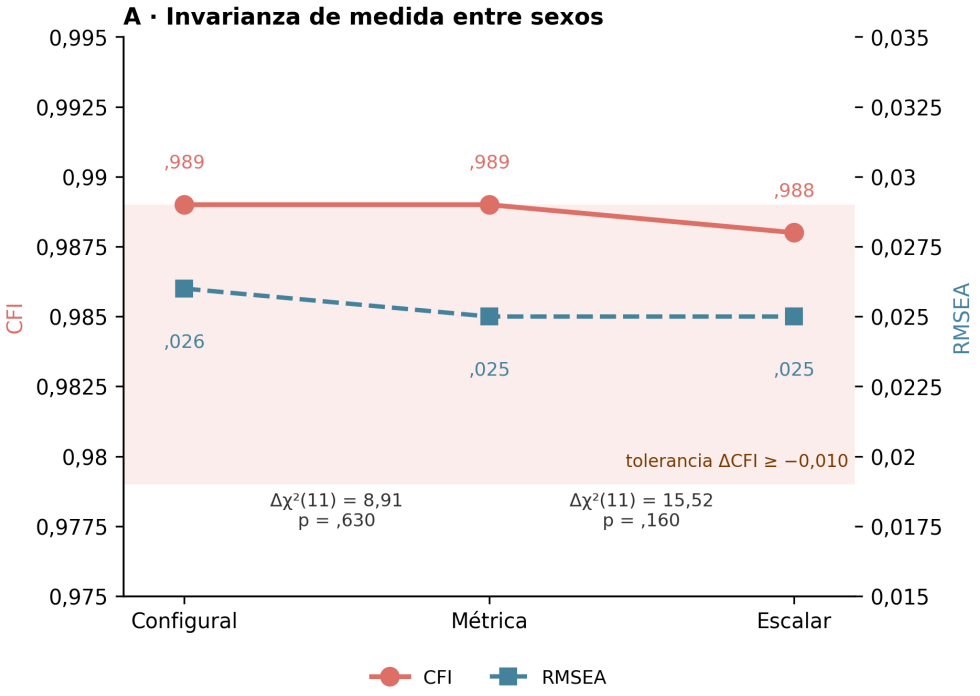
Trayectoria	β mujeres	β varones	$\Delta\beta$ (mujeres – varones)	$\Delta\chi^2(1)$	p	Decisión
EIC → CCG	0,554	0,371	0,183	6,95	0,008	Difiere
CCG → MET	0,434	0,426	0,008	0,06	0,800	No difiere
EIC → MET	0,019	0,204	-0,185	3,40	0,065	No difiere
MET → CCN	0,267	0,147	0,120	1,50	0,221	No difiere
CCG → CCN	0,221	0,108	0,113	0,88	0,348	No difiere
EIC → CCN	0,018	0,140	-0,122	1,77	0,184	No difiere

Trayectoria	β mujeres	β varones	Δβ (mujeres − varones)	Δχ²(1)	p	Decisión
Pretest → CCG	0,198	-0,025	0,222	8,84	0,003	Difiere
Pretest → MET	0,164	0,138	0,026	0,11	0,738	No difiere
Pretest → CCN	0,581	0,632	-0,051	0,79	0,374	No difiere

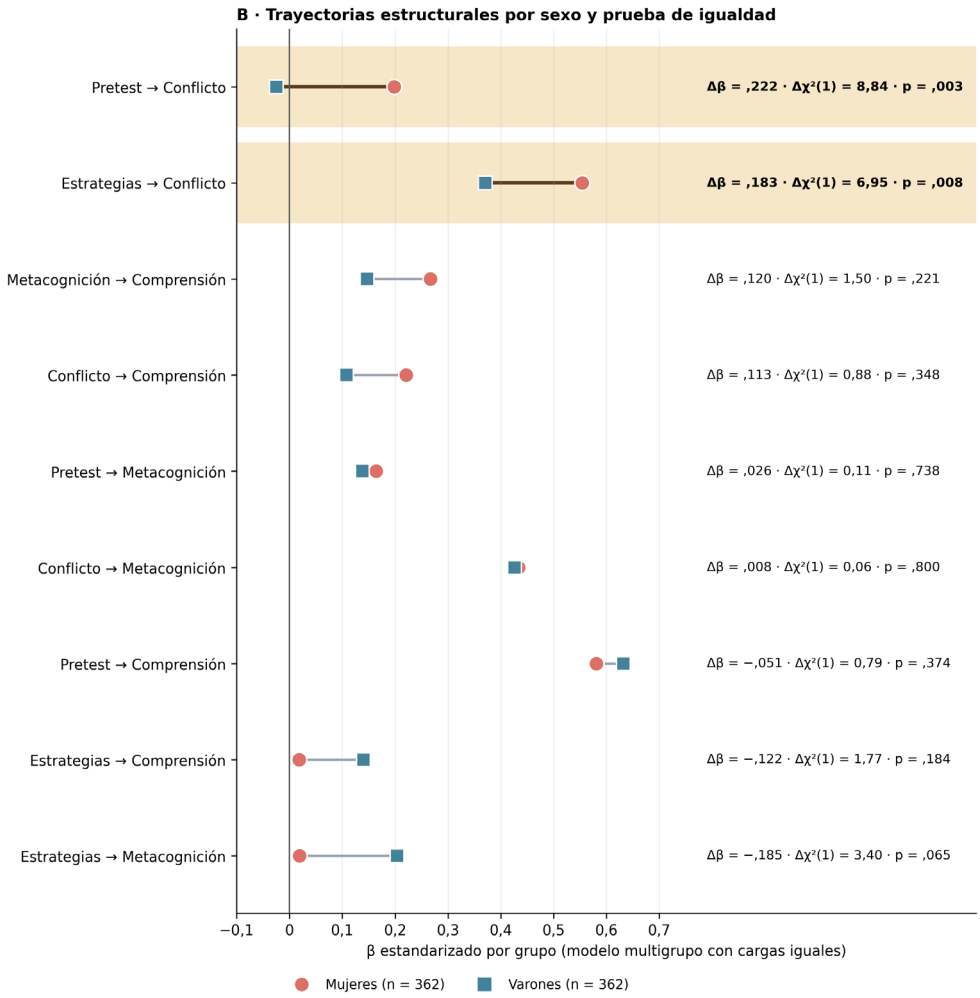
Nota: Coeficientes estandarizados del modelo multigrupo con cargas iguales (invarianza métrica). Δχ²(1) = cambio de χ² al fijar igual la trayectoria en los dos grupos. Pretest = comprensión conceptual inicial.

Figura 8-4

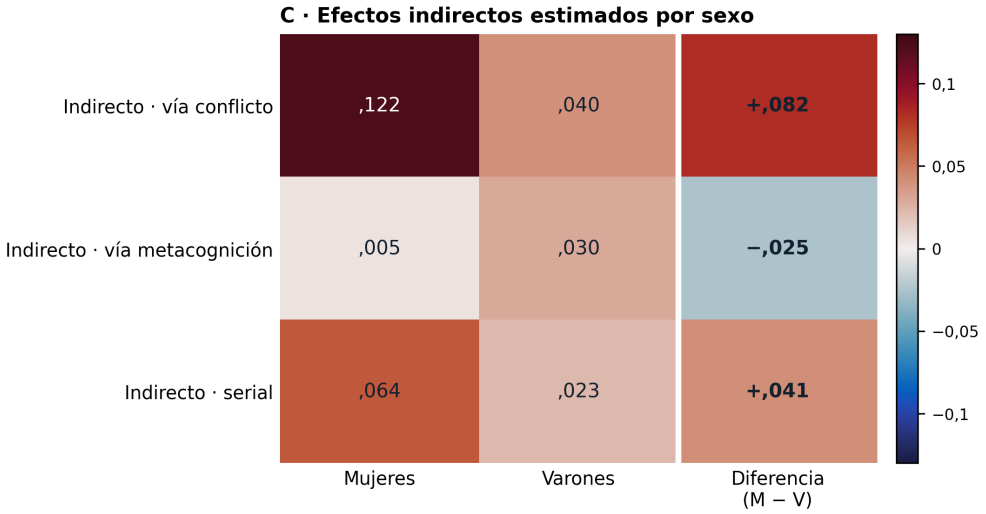
Invarianza de medida y diferencias por sexo en el modelo estructural



Nota. Mujeres, n = 362; varones, n = 362. Panel A: CFI y RMSEA de los modelos configural, métrico y escalar; la banda sombreada marca la tolerancia ΔCFI ≥ −0,010; métrico frente a configural, Δχ²(11) = 8,91, p = ,630; escalar frente a métrico, Δχ²(11) = 15,52, p = ,160.



Nota. Mujeres, n = 362; varones, n = 362. Panel B: coeficientes estandarizados por grupo en el modelo multigrupo con cargas iguales y prueba de igualdad de cada trayectoria ($\Delta\chi^2$ con un grado de libertad); en sombreado, las dos trayectorias que difieren: estrategias → conflicto cognitivo (,554 frente a ,371; $\Delta\chi^2(1) = 6{,}95$; p = ,008) y pretest → conflicto cognitivo (,198 frente a –,025; $\Delta\chi^2(1) = 8{,}84$; p = ,003).



Nota. Mujeres, n = 362; varones, n = 362. Panel C: efectos indirectos por sexo y su diferencia (mujeres menos varones): vía conflicto, ,122 frente a ,040 (+,082); vía metacognición, ,005 frente a ,030 (−,025); vía serial, ,064 frente a ,023 (+,041).

8.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre estrategias y comprensión, sino que el eslabón del mecanismo más intenso en las estudiantes no se tradujo en una reducción de la brecha conceptual: las mujeres mostraron un vínculo más fuerte entre las estrategias y el conflicto cognitivo, y solo en ellas la comprensión de partida se asoció con esa experiencia, pero su desventaja inicial persistió en el postest. Este resultado sugiere que un mecanismo más intenso no compensa por sí solo una desigualdad de partida, y obliga a leer el hallazgo de equidad con cautela. Desde una perspectiva crítica, el patrón es coherente con la idea de que el aprendizaje activo estrecha, pero no elimina, las brechas de rendimiento (Theobald et al., 2020), aunque aquí la evidencia procede de un diseño de dos ondas que no permite atribuir el cambio a la intervención.

El resultado central es que la asociación entre las estrategias y la comprensión es mayoritariamente indirecta: el 57,6 % del efecto total circula por el conflicto cognitivo y la regulación metacognitiva, y la vía serial es distinta de cero. Persiste un efecto directo pequeño, de modo que el

mecanismo propuesto explicaría la mayor parte, pero no la totalidad, de la relación. Esta lectura ofrece una explicación de proceso a la heterogeneidad que registran las síntesis de cambio conceptual (Aleknavičiūtė et al., 2023; Paçacı et al., 2024). Si el beneficio de las estrategias depende de que provoquen disonancia productiva, las aulas que adoptan formatos interactivos sin generar conflicto no obtendrían el efecto esperado; es precisamente el problema de implementación que describen Kera y Özdemir (2026) al documentar la variación entre 229 implementaciones de conflicto cognitivo en física. El modelo serial se ajustó mejor que el de mediadores paralelos, y la regulación metacognitiva fue el predictor proximal de mayor contribución única, lo que es compatible con la tesis de que el conflicto solo rinde cuando el estudiante lo convierte en supervisión de su propio razonamiento. Conviene subrayar que el orden entre ambos mediadores descansa en la teoría y en la comparación de modelos, no en la secuencia temporal de la medida.

En contraste con buena parte de la literatura, que compara métodos de enseñanza y atribuye las diferencias al formato, los resultados sugieren que el formato es un vehículo y no el ingrediente activo. Sundstrom et al. (2026) comparan los beneficios relativos de distintos métodos de aprendizaje activo para el aprendizaje conceptual de la física, y las diferencias entre métodos que ellos documentan resultan interpretables si cada método genera conflicto y trabajo metacognitivo en grados distintos. En la misma línea, la eficacia de la instrucción por pares se ha atribuido a los procesos cognitivos que activa y a las reglas con que se aplica (Gjerde y Hagane, 2024), el encuadre que el docente da a la participación condiciona el compromiso y las ganancias del estudiante (Kashyap et al., 2026) y los entornos de juego integrados con el contenido de la mecánica newtoniana plantean la misma pregunta sobre qué procesos median el aprendizaje (van der Linden et al., 2024). Nuestro modelo traduce esas observaciones a una hipótesis estructural contrastable: lo que se transmite es la experiencia de conflicto y su regulación, no la etiqueta de la estrategia. Aun así, las estrategias se midieron por percepción y no por observación de aula, por

lo que no puede descartarse que parte de la asociación refleje cómo los estudiantes que aprenden más perciben la enseñanza.

La dimensión de equidad merece un tratamiento más matizado. La brecha inicial en contra de las mujeres coincide en dirección con la descrita antes de la instrucción en física (Burkholder y Salehi, 2022) y con las diferencias por sexo en el inventario en estudiantes turcos, que además involucran la confianza y la calibración (Kaltakci y Özmen, 2026). Lo que aporta este estudio es un mecanismo: la confrontación de predicciones en grupo puede ofrecer un contexto de menor amenaza para revisar las propias ideas, y esa condición favorecería sobre todo a quienes llegan con menor comprensión inicial. Es una conjetura, no un resultado, pues el estudio no midió ansiedad ni percepción de amenaza. Resulta pertinente la evidencia de 70 países según la cual las autopercepciones operan como mecanismos de desigualdad de logro (Hofer et al., 2024); si el conflicto cognitivo es una experiencia vinculada con la autopercepción, la mayor intensidad observada en las mujeres podría reflejar esa vía, aunque esta interpretación excede los datos disponibles. Conviene además registrar que la vía metacognitiva directa desde las estrategias fue menor en las mujeres que en los varones, sin alcanzar significación, de modo que el beneficio diferencial se concentra en el primer eslabón y no en todo el mecanismo.

El contexto amplía el dominio de validez de la teoría, con las salvedades debidas. La evidencia sobre compromiso interactivo procede sobre todo de cursos universitarios; aquí el mecanismo se observa en secundaria pública, con ganancias normalizadas modestas y una población lingüísticamente diversa. Las condiciones de trabajo docente se asocian con el logro en el Perú (Rentería y Solano, 2025), y las brechas territoriales de desempeño en la región son amplias, como muestra el caso colombiano (Rodríguez et al., 2024). Que la estructura del mecanismo se sostenga en un entorno tan distinto de aquel en el que se formuló sugiere que podría tratarse de una propiedad general del aprendizaje conceptual de la física, aunque la muestra de seis distritos de una sola área metropolitana impide generalizar a la población escolar peruana. En el mismo sentido, las ideas erróneas sobre

fuerza y movimiento persisten en estudiantes de contextos muy diversos (Bouzid et al., 2025), y el efecto de los modelos de cambio conceptual sobre el logro y la autoeficacia ha sido documentado en la enseñanza de la física (Ugwuanyi et al., 2023), lo que apunta a que el mecanismo no sería exclusivo del contexto estudiado.

Las implicaciones prácticas se derivan de lo anterior con sobriedad. No basta con prescribir formatos interactivos: lo que transmite el beneficio es la experiencia de conflicto que esos formatos provocan y el trabajo metacognitivo que la sigue. Las prácticas que el modelo vincula con el mecanismo —pedir predicciones explícitas, contrastarlas con demostraciones o simulaciones, discutirlas entre pares y usar las respuestas para ajustar la explicación— son de bajo costo y compatibles con las condiciones de las escuelas públicas periurbanas, como muestran secuencias contextualizadas de mecánica newtoniana con enfoque STEM (Lucero et al., 2026) y estrategias apoyadas en software de simulación para movimiento y fuerza (Cevallos y Mestre, 2023). La correlación intraclase de la percepción de las estrategias indica que se trata de una práctica de aula y, por tanto, de un objetivo pertinente para el acompañamiento docente. Que el primer eslabón opere con más fuerza en las mujeres sugiere que estas prácticas pueden servir a la equidad sin intervenciones diferenciadas, aunque la persistencia de la brecha advierte contra esperar que cierren por sí solas la desventaja de partida.

Cinco limitaciones acotan las conclusiones. La muestra es no probabilística y procede de seis distritos de una sola área metropolitana. Las tres percepciones se midieron en el mismo momento, por lo que el orden entre el conflicto cognitivo y la regulación metacognitiva descansa en la teoría. Las estrategias se midieron por percepción de los estudiantes. La comprensión conceptual, medida con tres indicadores de contenido, obtuvo una AVE de 0,427, inferior a 0,50, aunque con fiabilidad compuesta de 0,689 y validez discriminante adecuada. Por último, la ganancia normalizada tiene limitaciones conocidas como métrica de comparación, por lo que se complementó con d_z. Las líneas futuras incluyen diseños de tres o más ondas que midan los mediadores entre pretest y postest, la

observación sistemática de las prácticas de aula, modelos multinivel con un número de aulas suficiente para estimar el mecanismo en ese nivel y la incorporación de modelos de diagnóstico cognitivo para identificar qué concepciones cambian.

8.6 CONCLUSIONES

Las estrategias didácticas de compromiso interactivo se asocian con la comprensión de la mecánica newtoniana sobre todo de forma indirecta: provocan conflicto cognitivo, el conflicto activa la regulación metacognitiva y ambos procesos se vinculan con el cambio conceptual al cierre de la unidad, con un 57,6 % del efecto total transmitido por esas vías. El modelo serial superó a las alternativas sin mediación, con mediación completa y con mediadores paralelos, lo que respalda un mecanismo de dos eslabones en el que la regulación metacognitiva es el predictor proximal de mayor contribución única.

La medida resultó equivalente entre mujeres y varones, de modo que la comparación entre sexos es válida. El primer eslabón del mecanismo, la capacidad de las estrategias para provocar conflicto cognitivo, opera con más fuerza en las estudiantes, que parten de una comprensión inicial más baja. Esa mayor intensidad no cerró la brecha conceptual al término de la unidad, por lo que debe entenderse como una condición favorable y no como una compensación suficiente.

En contextos periurbanos con ganancias modestas, el valor de las estrategias interactivas depende de su capacidad para producir disonancia productiva y trabajo metacognitivo, no del formato en sí. Para la formación y el acompañamiento docente, ello significa diseñar actividades que obliguen a predecir, contrastar con evidencia y discutir, y verificar que el conflicto se convierta en revisión explícita del razonamiento. Esa capacidad es, además, una palanca de equidad de género en la enseñanza de la física. Los resultados describen asociaciones en una muestra no probabilística y no autorizan inferencias causales ni generalizaciones a la población escolar,

pero delimitan con claridad qué procesos conviene medir en la próxima generación de estudios longitudinales y observacionales.

8.7 REFERENCIAS BIBLIOGRÁFICAS

Aguirre, J., & Aguirre, L. (2025). Cinemática bidimensional en estudiantes de nivelación: análisis de concepciones alternativas y perspectivas didácticas. *Reincisol., 4*(7), 1734-1758. https://doi.org/10.59282/reincisol.v4(7)1734-1758

Aguirre, J., & López, M. (2026). Estrategias de aprendizaje activo en estudiantes de secundaria. Revisión sistemática. *Cienciamatria, 12*(22), 4-17. https://doi.org/10.35381/cm.v12i22.1970

Aleknavičiūtė, V., Lehtinen, E., & Södervik, I. (2023). Thirty years of conceptual change research in biology – A review and meta-analysis of intervention studies. *Educational Research Review, 41*, 100556. https://doi.org/10.1016/j.edurev.2023.100556

Bazán, A., Hernández, E., Bazán, W., & Tresierra, M. (2022). Effects of Opportunities to Learn on Peruvian Students' Science Achievement in Program for International Student Assessment 2015. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.897473

Bouzid, T., Darhmaoui, H., & Kaddari, F. (2025). Force and motion misconceptions in Moroccan high school science majors: insights from video game activity. *Research in Science & Technological Education, 43*(4), 1247-1268. https://doi.org/10.1080/02635143.2024.2428618

Buncher, J., Nissen, J., Van Dusen, B., & Talbot, R. (2025). Is the Force Concept Inventory biased across the intersections of gender and

race?. *Physical Review Physics Education Research, 21*(1). https://doi.org/10.1103/physrevphyseducres.21.010137

Burkholder, E., & Salehi, S. (2022). Exploring the pre-instruction gender gap in physics. *Plos One, 17*(7), e0271184. https://doi.org/10.1371/journal.pone.0271184

Cevallos, E., & Mestre, U. (2023). Estrategia didáctica para el uso del software GeoGebra en el aprendizaje del movimiento y la fuerza en los estudiantes de Bachillerato General Unificado. *Educação Matemática Debate, 7*(13), 1-24. https://doi.org/10.46551/emd.v7n13a10

Christman, E., Miller, P., & Stewart, J. (2024). Beyond normalized gain: Improved comparison of physics educational outcomes. *Physical Review Physics Education Research, 20*(1). https://doi.org/10.1103/physrevphyseducres.20.010123

Coletta, V. (2023). Evidence for a normal distribution of normalized gains. *Physical Review Physics Education Research, 19*(1). https://doi.org/10.1103/physrevphyseducres.19.010111

Fleur, D., Bredeweg, B., & van den Bos, W. (2021). Metacognition: ideas and insights from neuro- and educational sciences. *npj Science of Learning, 6*(1). https://doi.org/10.1038/s41539-021-00089-5

Gjerde, V., & Hagane, S. (2024). Enhancing peer instruction in physics: Understanding cognitive processes and refining rules. *Physical Review Physics Education Research, 20*(1). https://doi.org/10.1103/physrevphyseducres.20.010134

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74. https://doi.org/10.1119/1.18809

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hernández, C., Prada, R., & Álvarez, F. (2023). Evaluation of Newton's Law Learning. *Scientia et Technica*, *28*(03), 144-149. https://doi.org/10.22517/23447214.24530

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141-158. https://doi.org/10.1119/1.2343497

Hofer, S., Heine, J., Besharati, S., Yip, J., Reinhold, F., & Brummelman, E. (2024). Self-perceptions as mechanisms of achievement inequality: evidence across 70 countries. *npj Science of Learning*, *9*(1). https://doi.org/10.1038/s41539-023-00211-9

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kaltakci, D., & Özmen, K. (2026). Examining Gender Differences in the Force Concept Inventory (FCI) in a Turkish Context: Accuracy,

Confidence and Bias Score Comparisons. *Social Sciences, 15*(3), 164. https://doi.org/10.3390/socsci15030164

Kashyap, J., Devaty, R., & Singh, C. (2026). Instructor Framing and Incentives Shape Physics Students' Engagement and Learning Gains from an Inquiry-Based Electrostatics Tutorial on the Method of Images. *Education Sciences, 16*(4), 594. https://doi.org/10.3390/educsci16040594

Kera, B., & Özdemir, O. (2026). Beyond uniformity: variations in the implementation of cognitive conflict strategies in physics education. *Physics Education, 61*(4), 045040. https://doi.org/10.1088/1361-6552/ae810f

Kersting, M., Karlsen, S., Ødegaard, M., Olufsen, M., Kjærnsli, M., & Suhr, M. (2023). Studying the quality of inquiry-based teaching in science classrooms. A systematic video study of inquiry-based science teaching in primary and lower-secondary schools. *International Journal of Science Education, 45*(17), 1463-1484. https://doi.org/10.1080/09500693.2023.2213386

Le, V., Nissen, J., Tang, X., Zhang, Y., Mehrabi, A., Morphew, J., Chang, H., & Van Dusen, B. (2025). Applying cognitive diagnostic models to mechanics concept inventories. *Physical Review Physics Education Research, 21*(1). https://doi.org/10.1103/physrevphyseducres.21.010103

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development,

methodological challenges, state of the art, and future perspectives. *Social Science Research, 110,* 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Li, Y., & Singh, C. (2023). Considering Statistically Equivalent Models when using Structural Equation Modeling: an Example from Physics Identity. *International Journal of Innovation in Science and Mathematics Education, 31*(4). https://doi.org/10.30722/ijisme.31.04.001

Lucero, A., Villegas, M., Benegas, J., & Hernández, C. (2026). Aprendizaje conceptual de la mecánica newtoniana en el contexto de la educación vial: un proyecto interdisciplinario con enfoque STEM. *Revista de Enseñanza de la Física, 38*(1), 155-163. https://doi.org/10.55767/2451.6007.v38.n1.52372

Maassen, E., D'Urso, E., van Assen, M., Nuijten, M., De Roover, K., & Wicherts, J. (2025). The dire disregard of measurement invariance testing in psychological science. *Psychological Methods, 30*(5), 966-979. https://doi.org/10.1037/met0000624

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Morley, A., Nissen, J., & Van Dusen, B. (2023). Measurement invariance across race and gender for the Force Concept Inventory. *Physical Review Physics Education Research, 19*(2). https://doi.org/10.1103/physrevphyseducres.19.020102

Moshagen, M., & Bader, M. (2024). semPower: General power analysis for structural equation models. *Behavior Research Methods, 56*(4), 2901-2922. https://doi.org/10.3758/s13428-023-02254-7

Navarro, M., Padilla, J., & Benítez, I. (2024). Analyzing Measurement Invariance for Studying the Gender Gap in Educational Testing. *European Journal of Psychological Assessment, 40*(5), 412-426. https://doi.org/10.1027/1015-5759/a000820

Öztürk, B., Kaya, M., & Demir, M. (2022). Does inquiry-based learning model improve learning outcomes? A second-order meta-analysis. *Journal of Pedagogical Research.* https://doi.org/10.33902/jpr.202217481

Paçacı, C., Üstün, U., & Özdemir, O. (2024). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching, 61*(6), 1263-1325. https://doi.org/10.1002/tea.21887

Podsakoff, P., Podsakoff, N., Williams, L., Huang, C., & Yang, J. (2024). Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix. *Annual Review of Organizational Psychology and Organizational Behavior, 11*(1), 17-61. https://doi.org/10.1146/annurev-orgpsych-110721-040030

Rentería, J., & Solano, D. (2025). The parabolic path of teacher well-being and student learning achievement in Peru. *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-05736-5

Rodríguez, W., Rodríguez, E., & Gómez, J. (2024). The Rural-Urban Student Performance Gap in Colombia. *Desarrollo y Sociedad*(97), 59-75. https://doi.org/10.13043/dys.97.3

Sundstrom, M., Gambrell, J., Green, C., Traxler, A., & Brewe, E. (2026). Relative benefits of different active learning methods to conceptual physics learning. *Nature Physics*, *22*(7), 1031-1036. https://doi.org/10.1038/s41567-026-03307-2

Theobald, E., Hill, M., Tran, E., Agrawal, S., Arroyo, E., Behling, S., Chambwe, N., Cintrón, D., Cooper, J., Dunster, G., Grummer, J., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M., Littlefield, C., Lowe, A., Newman, S., Okolo, V., Olroyd, S., Peecook, B., Pickett, S., Slager, D., Caviedes, I., Stanchak, K., Sundaravardan, V., Valdebenito, C., Williams, C., Zinsli, K., & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476-6483. https://doi.org/10.1073/pnas.1916903117

Ugwuanyi, C., Ezema, M., & Orji, E. (2023). Evaluating the Instructional Efficacies of Conceptual Change Models on Students' Conceptual Change Achievement and Self-Efficacy in Particulate Nature Matter in Physics. *Sage Open*, *13*(1). https://doi.org/10.1177/21582440231153851

Ulu, Y., & Yerdelen, S. (2024). Metacognition and epistemic cognition in physics are related to physics identity through the mediation of physics self-efficacy. *Physical Review Physics Education Research*, *20*(1). https://doi.org/10.1103/physrevphyseducres.20.010130

van der Linden, A., Meulenbroeks, R., & van Joolingen, W. (2024). Learning Newtonian mechanics with an intrinsically integrated educational game. *Journal of Computer Assisted Learning*, *40*(4), 1500-1510. https://doi.org/10.1111/jcal.12966

Wang, Y., & Rhemtulla, M. (2021). Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial. Advances in Methods and Practices in Psychological Science, 4(1). https://doi.org/10.1177/2515245920918253

CAPÍTULO IX

HERRAMIENTAS DIGITALES Y RAZONAMIENTO LÓGICO-MATEMÁTICO EN LA EDUCACIÓN PRIMARIA: EVIDENCIA CUASIEXPERIMENTAL CON INFERENCIA ROBUSTA POR AULA

Lina Lucy Fabian Ojeda

Universidad Nacional de Educación Enrique Guzmán y Valle

Karla Sofia Dongo Ortiz

Universidad Nacional de Educación Enrique Guzmán y Valle

Fiorella Ñahuin Antonio

Universidad Nacional de Educación Enrique Guzmán y Valle

Resumen

Este estudio estimó el efecto de una intervención de doce semanas con herramientas digitales sobre el razonamiento lógico-matemático de estudiantes de quinto grado de primaria y sobre sus tres componentes: clasificación y seriación, patrones y regularidades, y resolución de problemas. Se realizó un cuasiexperimento pretest-postest con grupo de comparación no equivalente en 24 aulas intactas de seis instituciones públicas de Lima (Perú), con 543 estudiantes (272 experimentales y 271 de comparación) durante 2025. El tamaño de la muestra se calculó a priori para el contraste más exigente y se verificó por simulación con el análisis declarado. El efecto se estimó con análisis de covarianza del postest controlando el pretest, errores estándar robustos por aula e inferencia con 22 grados de libertad, y se contrastó con modelos multinivel, bootstrap salvaje por conglomerado, análisis por rangos y exclusión de instituciones. La intervención mejoró el puntaje total ajustado en 4.42 puntos (IC 95 % [1.80, 7.03]; g = 0.47) y cada dimensión tras el ajuste por multiplicidad (g entre 0.28 y 0.36). La proporción de estudiantes en el nivel satisfactorio pasó de 11.8 % a 30.5 % en la condición experimental y de 14.8 % a 18.8 % en la de comparación, y las estimaciones se mantuvieron con todos los estimadores alternativos. El estudio desplaza la pregunta del acceso a la tecnología hacia su integración didáctica en tareas de razonamiento, verifica la estructura factorial del instrumento en ambas mediciones y trata las aulas como unidad de asignación en el cálculo de la muestra y en la inferencia.

Palabras clave: herramientas digitales; razonamiento lógico-matemático; educación primaria; cuasiexperimento; análisis de covarianza; diseño por conglomerados.

Abstract

This study estimated the effect of a twelve-week digital-tools intervention on fifth-grade students' logical-mathematical reasoning and on its three components:

classification and seriation, patterns and regularities, and problem solving. A pretest-posttest quasi-experiment with a nonequivalent comparison group was conducted in 24 intact classrooms from six public schools in Lima (Peru), with 543 students (272 experimental, 271 comparison) during 2025. Sample size was computed a priori for the most demanding contrast and verified by simulation under the declared analysis. The effect was estimated with analysis of covariance of the posttest adjusting for the pretest, classroom-robust standard errors and inference on 22 degrees of freedom, and was checked against multilevel models, wild cluster bootstrap, rank-based analysis and leave-one-school-out estimation. The intervention increased the adjusted total score by 4.42 points (95% CI [1.80, 7.03]; g = 0.47) and each of the three dimensions after multiplicity adjustment (g between 0.28 and 0.36). The share of students at the satisfactory level rose from 11.8% to 30.5% in the experimental condition and from 14.8% to 18.8% in the comparison condition, and estimates held under every alternative estimator. The study shifts the question from access to technology to its instructional integration in reasoning tasks, verifies the factor structure of the instrument at both time points, and treats classrooms as the unit of allocation in both sample-size planning and inference.

Keywords: digital tools; logical-mathematical reasoning; primary education; quasi-experiment; analysis of covariance; cluster design.

9.1 INTRODUCCIÓN

El razonamiento lógico-matemático —la capacidad de clasificar y ordenar objetos según relaciones, reconocer y generalizar regularidades y encadenar inferencias para resolver problemas— sostiene el aprendizaje de la matemática escolar. Las habilidades lógicas de los niños predicen su rendimiento matemático posterior más allá de la memoria de trabajo, y entrenarlas produce avances en matemática (Nunes et al., 2007); en la preadolescencia, el razonamiento transitivo se asocia con ese rendimiento por encima de los factores cognitivos generales (Wong y Morsanyi, 2023), y el razonamiento inductivo se vincula con el logro en las áreas STEM (Van Vo y Csapó, 2023). Se trata de una competencia que sigue

desarrollándose durante la secundaria (Öz y Işık, 2024) y que se resiente cuando la escolarización se interrumpe, como mostró el cierre de escuelas por la pandemia en la educación primaria (Coskun y Kara, 2022).

Las herramientas digitales promueven el fortalecimiento de estos procesos, con efectos positivos pero de magnitud muy variable. La tecnología en la enseñanza matemática de primaria alcanzó un efecto medio de 0.483 (Akçay et al., 2021), y en secundaria las herramientas digitales de matemática y ciencias registraron un g = 0.65, acrecentado ante la formación del profesorado (Hillmayr et al., 2020). Los juegos digitales benefician a las áreas STEM de primaria (Behnamnia et al., 2024), mientras que plataformas lúdicas como Kahoot! poseen síntesis propias (Özdemir, 2025). La inconsistencia de los estudios primarios sobre resolución de problemas motivó un metaanálisis de tres niveles (Cai et al., 2025), y los juegos educativos para Android ejercen un efecto moderado sobre el razonamiento matemático (Pratama, 2025). Los mecanismos plausibles varían: manipulativos virtuales y concretos eficaces para la enseñanza de fracciones en cuarto y quinto grado (Vessonen et al., 2021) y evaluados en quinto grado según el nivel de logro (Ahmad y Siller, 2024); geometría con soporte digital (Scippo et al., 2025); diseño lúdico que conecta representaciones (Debrenti, 2024) junto al andamiaje percibido por los estudiantes en los juegos (Sun et al., 2021); y sistemas de tutoría que adaptan la tarea (Son, 2024; Schaaf et al., 2026). La evidencia discrepa: un juego de tablero numérico igualó a la enseñanza habitual (James et al., 2024) y los recursos no digitales también potencian la matemática en primaria y secundaria (Tripo y Debrenti, 2026).

Tres limitaciones de esta evidencia motivan el estudio. La primera es de constructo: gran parte de los trabajos mide el rendimiento curricular y no el razonamiento que lo sustenta. Los sistemas de tutoría inteligente se han evaluado sobre todo por el desempeño dentro del propio sistema, como en la evaluación longitudinal realizada durante el cierre de escuelas (Spitzer y Moeller, 2024); la integración del pensamiento computacional en la matemática de primaria ha avanzado con más descripción de actividades que medición de procesos de razonamiento (Nordby et

al., 2022; Cui y Ng, 2021), aunque ya existen programas evaluados con pruebas validadas (Zapata et al., 2024). La segunda es de diseño: abundan los diseños de un solo grupo con pretest y postest (Mamolo y Sugano, 2024), y los cuasiexperimentos asignan aulas intactas pero analizan a los estudiantes como si lo hubieran sido uno por uno, lo que subestima los errores estándar cuando las puntuaciones de una misma aula están correlacionadas (Hedges y Hedberg, 2007; Huang y Li, 2021); los ensayos por conglomerados en primaria siguen siendo escasos (Yihan et al., 2026; Charalambous y Charalambous, 2025). La tercera es de contexto: la evidencia procede mayoritariamente de sistemas con alta disponibilidad tecnológica, mientras que en América Latina las brechas digitales, incluida la de género, condicionan el uso pedagógico de la tecnología (Ancheta et al., 2021), y en el Perú el bajo desempeño en matemática ha motivado el análisis de sus predictores afectivos y socioeconómicos (Santaria y Mamani, 2026).

El acceso, sin embargo, no basta: la dotación de computadoras en las escuelas rurales del Perú se ha evaluado incluso en el largo plazo (Cueto et al., 2025), y la pregunta pertinente es si una secuencia didáctica que use herramientas digitales para trabajar explícitamente los procesos de razonamiento mejora esos procesos. Este estudio la responde con un diseño que trata el aula como unidad de asignación y con una medición cuya estructura se verifica antes de comparar condiciones. Su aporte consiste en ofrecer una estimación con inferencia robusta por conglomerados de un efecto que la literatura suele reportar sin esa corrección, y en separar el efecto sobre cada componente del razonamiento, información que los metaanálisis no pueden recuperar porque los estudios primarios no la reportan.

9.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

9.2.1 El razonamiento lógico-matemático y sus componentes

Desde una perspectiva constructivista, el razonamiento lógico-matemático se desarrolla a partir de la acción del niño sobre los objetos

y de la reflexión sobre las relaciones que establece entre ellos. En quinto grado, con estudiantes de diez a once años, tres componentes concentran las demandas del currículo. La clasificación y la seriación implican agrupar por atributos y ordenar según una relación, operaciones sobre las que descansan la inclusión de clases y la comprensión del sistema de numeración, y cuyo papel en el aprendizaje matemático está documentado (Nunes et al., 2007; Wong y Morsanyi, 2023). Los patrones y regularidades suponen reconocer una estructura que se repite o crece y expresarla como regla general, apoyándose en representaciones visuales y ejemplos reales (Rizos y Gkrekas, 2024). La resolución de problemas integra ambos componentes en la selección de datos, la planificación de pasos y la verificación de resultados, y su desarrollo en los últimos grados de primaria ha sido objeto de ensayos por conglomerados (Charalambous y Charalambous, 2025). El razonamiento inductivo atraviesa los tres componentes (Van Vo y Csapó, 2023).

9.2.2 Vías por las que las herramientas digitales pueden mejorar el razonamiento

Las herramientas digitales pueden actuar por tres vías. La primera es la manipulación con retroalimentación inmediata: los manipulativos virtuales permiten clasificar, ordenar y transformar representaciones y señalan el error en el momento (Vessonen et al., 2021). La segunda es la adaptación de la dificultad, que mantiene la práctica en la zona en la que el razonamiento progresa (Son, 2024; Schaaf et al., 2026). La tercera es el andamiaje mediante el juego, percibido por los estudiantes de primaria (Sun et al., 2021), sensible al número de intentos que el juego permite (Cubillos et al., 2023) y favorable a la conexión entre representaciones (Debrenti, 2024). Estas vías no garantizan el efecto: la calidad educativa de muchas aplicaciones comerciales es baja (Meyer et al., 2021), un juego de tablero numérico no superó a la enseñanza habitual (James et al., 2024) y las intervenciones no digitales con material físico también desarrollan la resolución de problemas (Tripo y Debrenti, 2026). Por ello, la intervención alineó las tres vías con los componentes del razonamiento

y la comparación se estableció contra una enseñanza habitual que trabajó las mismas competencias durante el mismo tiempo lectivo, de modo que el contraste aísla el medio y la retroalimentación y no la cantidad de instrucción.

9.2.3 Hipótesis

- **H1.** Al controlar el razonamiento inicial, los estudiantes de la condición experimental obtienen un puntaje total de razonamiento lógico-matemático mayor en el postest que los de la condición de comparación.
- **H2a.** El efecto de H1 se observa en la clasificación y la seriación.
- **H2b.** El efecto de H1 se observa en los patrones y regularidades.
- **H2c.** El efecto de H1 se observa en la resolución de problemas.

9.3 MATERIALES Y MÉTODOS

9.3.1 Diseño y participantes

Se empleó un cuasiexperimento pretest-postest con grupo de comparación no equivalente, entre abril y julio de 2025, en seis instituciones públicas de Lima Metropolitana (Perú). En cada institución, dos de las cuatro secciones de quinto grado se asignaron a la condición experimental y dos a la de comparación según la disponibilidad del aula de innovación y de los horarios; la asignación fue, por tanto, por aula intacta y dentro de cada institución, con 24 aulas como unidad de asignación. Las instituciones se seleccionaron por conveniencia, de modo que los resultados describen a las aulas analizadas. Participaron los estudiantes que rindieron ambas mediciones; las aulas tuvieron entre 19 y 27 estudiantes (mediana = 22.5).

9.3.2 Tamaño de la muestra

El tamaño se calculó a priori para el contraste más exigente, cada dimensión con ajuste de Holm. Se adoptó un efecto de planificación d = 0.40, inferior al efecto medio de la tecnología en la matemática de primaria (0.483; Akçay

et al., 2021) y situado entre el de las herramientas de seguimiento docente (0.10; Faber et al., 2023) y el de las herramientas digitales en secundaria (0.65; Hillmayr et al., 2020), con síntesis de rendimiento matemático porque los efectos de referencia dependen del tipo de resultado (Wilson et al., 2026), una correlación pretest-postest $\rho = .60$ y potencia de .80. Resultaron 64 estudiantes por grupo para el total y 86 por dimensión ($\alpha = .0167$). Con asignación por aula, el efecto de diseño ($m = 23$, CCI = .05; Hedges y Hedberg, 2007; Ouyang et al., 2023) fue 2.10, lo que exige 362 estudiantes, equivalentes a 8 aulas por condición. Una simulación de Monte Carlo con el análisis declarado (errores robustos por aula, $gl = J - 2$, 2000 réplicas) dio potencias por dimensión de .613, .752, .796 y .838 con 8, 10, 11 y 12 aulas por condición (error de simulación = .008); se diseñó entonces con 12 aulas por condición y la muestra analizada fue $N = 543$. Con $N = 362$, el análisis multivariado detecta con potencia .80 una distancia de Mahalanobis $D = 0.35$, y el análisis factorial confirmatorio de 18 ítems distingue un RMSEA de .05 de uno de .08 con potencia > .99 (MacCallum et al., 1996).

9.3.3 Intervención

La Tabla 9-1 resume las dos condiciones. La intervención comprendió 36 sesiones de 45 minutos, tres por semana, en el aula de innovación y dentro del horario de Matemática, con un componente del razonamiento priorizado en cada bloque de cuatro semanas. Las docentes experimentales recibieron ocho horas de formación y acompañamiento quincenal, y el uso de las herramientas se registró por sesión; las aulas de comparación trabajaron las mismas competencias con igual número de sesiones y minutos, sin dispositivos.

Tabla 9-1

Estructura de la intervención con herramientas digitales y de la enseñanza habitual del grupo de comparación

Semanas	Dimensión priorizada	Herramientas y tareas del grupo experimental	Enseñanza del grupo de comparación
1–4	Clasificación y seriación	Manipulativos virtuales para clasificar por uno y dos atributos, ordenar series crecientes y decrecientes y completar matrices lógicas, con retroalimentación inmediata	Fichas impresas y material concreto con las mismas competencias del área de Matemática
5–8	Patrones y regularidades	Secuencias numéricas y geométricas en un entorno de geometría dinámica; generalización verbal de la regla y prueba de conjeturas	Secuencias en el cuaderno de trabajo y en la pizarra
9–12	Resolución de problemas	Problemas aritméticos de varios pasos en una plataforma de práctica adaptativa que regula la dificultad según el desempeño; juegos digitales de lógica con niveles	Problemas del texto escolar resueltos en forma individual y revisados en plenaria
1–12	Transversal	Tres sesiones semanales de 45 minutos dentro del horario de Matemática (36 sesiones); cuaderno de registro de uso por aula	Mismo número de sesiones y minutos, sin dispositivos

Nota: Ambas condiciones trabajaron las mismas competencias del área de Matemática durante el mismo tiempo lectivo; la diferencia entre condiciones fue el medio y la retroalimentación de las tareas.

9.3.4 Instrumento y procedimiento

El razonamiento se midió con una prueba de 18 ítems de respuesta construida, seis por dimensión (clasificación y seriación, patrones y regularidades, resolución de problemas), elaborada a partir de tareas de razonamiento lógico, transitivo e inductivo (Nunes et al., 2007; Wong y Morsanyi, 2023; Van Vo y Csapó, 2023). Una rúbrica de crédito parcial (0 a 3) llevó cada dimensión a 0–18 puntos y el total a 0–54; tres especialistas en didáctica de la matemática revisaron los ítems. Los niveles de logro

fueron en inicio (≤ 21), en proceso (22–37) y satisfactorio (≥ 38). La prueba se aplicó en papel, en forma colectiva y con la misma forma, durante 60 minutos, la semana previa y la siguiente a la intervención; evaluadoras externas la administraron y evaluadoras ciegas a la condición la puntuaron.

9.3.5 Análisis estadístico

La estructura se examinó con un análisis factorial confirmatorio de tres factores correlacionados, por máxima verosimilitud en cada medición y frente a un modelo unidimensional; la fiabilidad se estimó con ω de McDonald y α de Cronbach, y la equivalencia inicial con t de Welch, χ^2 y g de Hedges. El efecto se estimó con ANCOVA del postest sobre la condición, con el pretest como covariable, para el total (H1) y cada dimensión (H2a–H2c), con errores estándar robustos por aula (CR1) e inferencia t con $J - 2 = 22$ grados de libertad (Huang y Li, 2021; Lee y Pustejovsky, 2024). Se añadieron un análisis multivariado (traza de Pillai), el ajuste de Holm, la verificación de supuestos y de la correlación intraclase residual, y los tamaños del efecto (η^2 parcial, g de Hedges ajustada, ganancia normalizada), interpretados con referencias empíricas de campo (Kraft, 2020). La robustez se examinó con un modelo multinivel, una ANCOVA con efectos fijos de institución, el bootstrap salvaje por conglomerado con pesos de seis puntos (Webb, 2023), el bootstrap de aulas, el análisis por rangos de Quade y la exclusión de cada institución. Los niveles de logro se compararon con Stuart-Maxwell y con χ^2 y V de Cramér. Se adoptó $\alpha = .05$ bilateral.

9.4 RESULTADOS

9.4.1 Participantes y equivalencia inicial

La muestra analizada comprendió 543 estudiantes en 24 aulas: 272 en las 12 aulas experimentales y 271 en las 12 de comparación (Tabla 9-2). Las condiciones no difirieron en la proporción de mujeres ni en la edad, y las diferencias iniciales de razonamiento fueron pequeñas y no significativas,

todas favorables a la condición de comparación, con la mayor en clasificación y seriación.

Tabla 9-2

Características de los participantes y equivalencia inicial entre condiciones (N = 543)

Variable		Experimental (n = 272)	Comparación (n = 272)	Prueba	p	g de Hedges
Mujeres, n (%)		145 (53.3)	140 (51.7)	$\chi^2(1) = 0.15$	.701	—
Edad (meses), M (DE)		127.81 (4.63)	127.20 (4.57)	$t(541.0) = 1.55$	.121	0.13
Pretest:	clasificación y seriación, M (DE)	8.26 (4.47)	8.88 (4.13)	$t(537.9) = -1.67$	.095	−0.14
	patrones y regularidades, M (DE)	9.15 (4.33)	9.39 (4.34)	$t(541.0) = -0.66$	.512	−0.06
	resolución de problemas, M (DE)	8.13 (4.28)	8.41 (4.38)	$t(540.6) = -0.75$	.456	−0.06
	puntaje total, M (DE)	25.54 (10.51)	26.68 (10.33)	$t(540.9) = -1.27$	.203	−0.11

Nota: 24 aulas de quinto grado de primaria en seis instituciones educativas públicas (dos aulas experimentales y dos de comparación por institución). Prueba t de Welch para las variables continuas; g de Hedges = diferencia estandarizada (experimental − comparación). Pretest: dimensiones de 0 a 18 puntos; puntaje total de 0 a 54.

9.4.2 Estructura y fiabilidad del instrumento

El modelo de tres factores reprodujo la matriz de correlaciones de los ítems en las dos mediciones (Tabla 9-3). Las cargas estandarizadas oscilaron entre .579 y .731 en el pretest y entre .579 y .745 en el postest.

Tabla 9-3

Cargas factoriales estandarizadas del análisis factorial confirmatorio de la prueba de razonamiento lógico-matemático en el pretest y el postest (N = 543)

Dimensión	Ítem	pretest		postest	
		λ pretest	EE	λ postest	EE
Clasificación y seriación	CS1	.637	—	.685	—
	CS2	.610	.081	.660	.069
	CS3	.643	.084	.636	.070
	CS4	.731	.085	.745	.072
	CS5	.633	.084	.579	.066
	CS6	.720	.084	.723	.066
Patrones y regularidades	PR1	.730	—	.743	—
	PR2	.582	.062	.590	.062
	PR3	.606	.065	.619	.065
	PR4	.676	.064	.640	.065
	PR5	.582	.064	.639	.061
	PR6	.686	.065	.675	.065
Resolución de problemas	RP1	.579	—	.583	—
	RP2	.660	.101	.739	.104
	RP3	.658	.098	.707	.104
	RP4	.655	.100	.660	.099
	RP5	.711	.096	.707	.101
	RP6	.703	.105	.736	.101

Nota: Estimación por máxima verosimilitud con las puntuaciones de crédito parcial (0–3) de cada ítem. λ = carga estandarizada; EE = error estándar de la estimación no estandarizada. El primer ítem de cada dimensión fija la escala del factor y no lleva error estándar. Ítems: CS = clasificación y seriación; PR = patrones y regularidades; RP = resolución de problemas.

Como detalla la Tabla 9-4, el ajuste del modelo de tres factores fue excelente en ambas mediciones, mientras que el modelo unidimensional ajustó peor ($\Delta\chi^2(3)$ con $p < .001$ en el pretest y el postest). La fiabilidad ω fue de .810 a .845 por dimensión, la varianza media extraída osciló entre .418 y .477 y las correlaciones entre factores fueron de .278 a .640 (Figura 9-1).

Tabla 9-4

Fiabilidad por dimensión y ajuste del modelo de medida de tres factores frente al modelo unidimensional

Indicador		Pretest	Postest
Clasificación y seriación:	ω de McDonald / α de Cronbach / AVE	.825 / .824 / .441	.832 / .831 / .454
Patrones y regularidades:		.810 / .809 / .418	.816 / .815 / .426
Resolución de problemas:		.824 / .823 / .439	.845 / .844 / .477
Puntaje total:	α de Cronbach	.885	.852
Tres factores:	χ^2(gl)	141.99 (132)	134.23 (132)
	CFI / TLI	.997 / .996	.999 / .999
	RMSEA [IC 90 %]	.012 [.000, .024]	.006 [.000, .022]
	SRMR	.028	.029
Unidimensional:	χ^2(gl)	896.93 (135)	1530.73 (135)
	CFI / TLI	.753 / .721	.554 / .495
	RMSEA [IC 90 %]	.102 [.096, .108]	.138 [.132, .144]
	SRMR	.085	.131

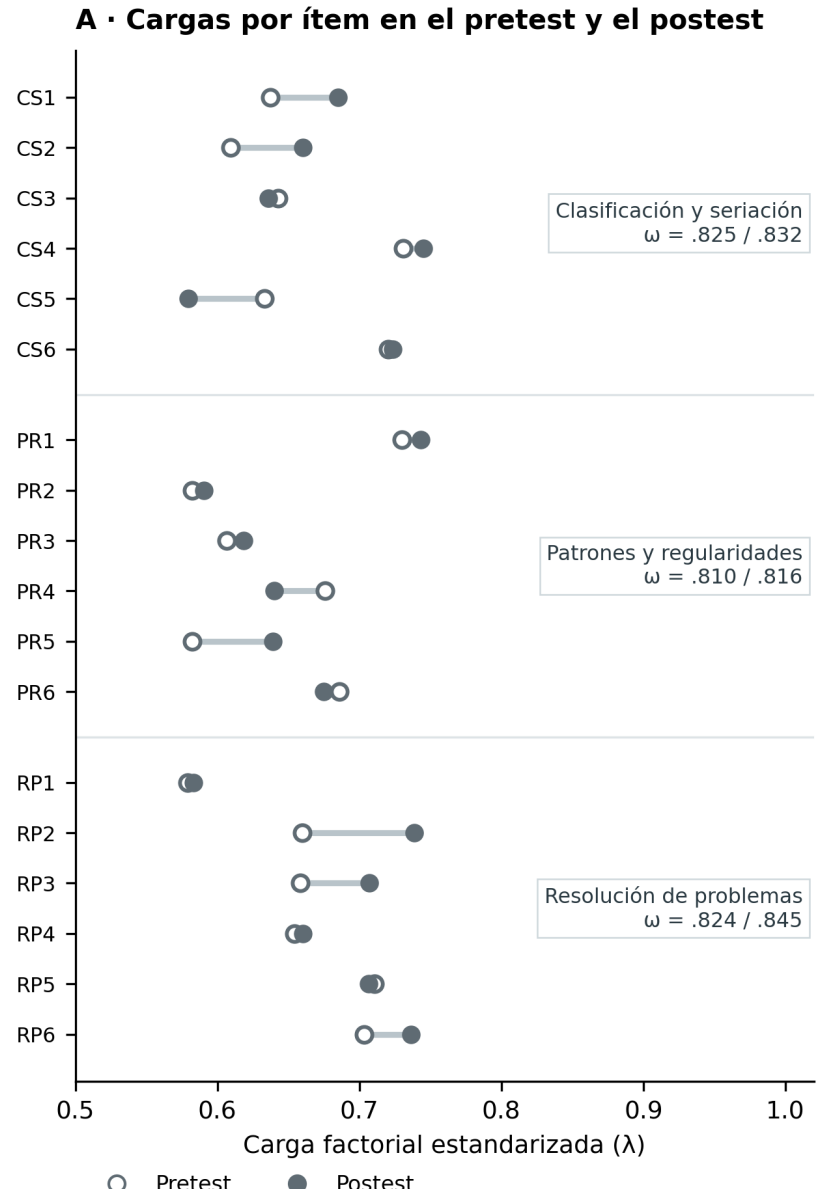
Nota: $\omega = (\Sigma\lambda)^2/[(\Sigma\lambda)^2 + \Sigma(1 - \lambda^2)]$ con las cargas estandarizadas de la Tabla 9-3; AVE = varianza media extraída. Correlaciones entre factores en el pretest: CS–PR = .569, CS–RP = .505, PR–RP = .640. Diferencia de χ^2 frente al modelo unidimensional: pretest $\Delta\chi^2(3) = 754.94$, $p < .001$; postest $\Delta\chi^2(3) = 1396.50$, $p < .001$. SRMR calculado sobre la matriz de correlaciones reproducida por el modelo.

9.4.3 Estadísticos descriptivos

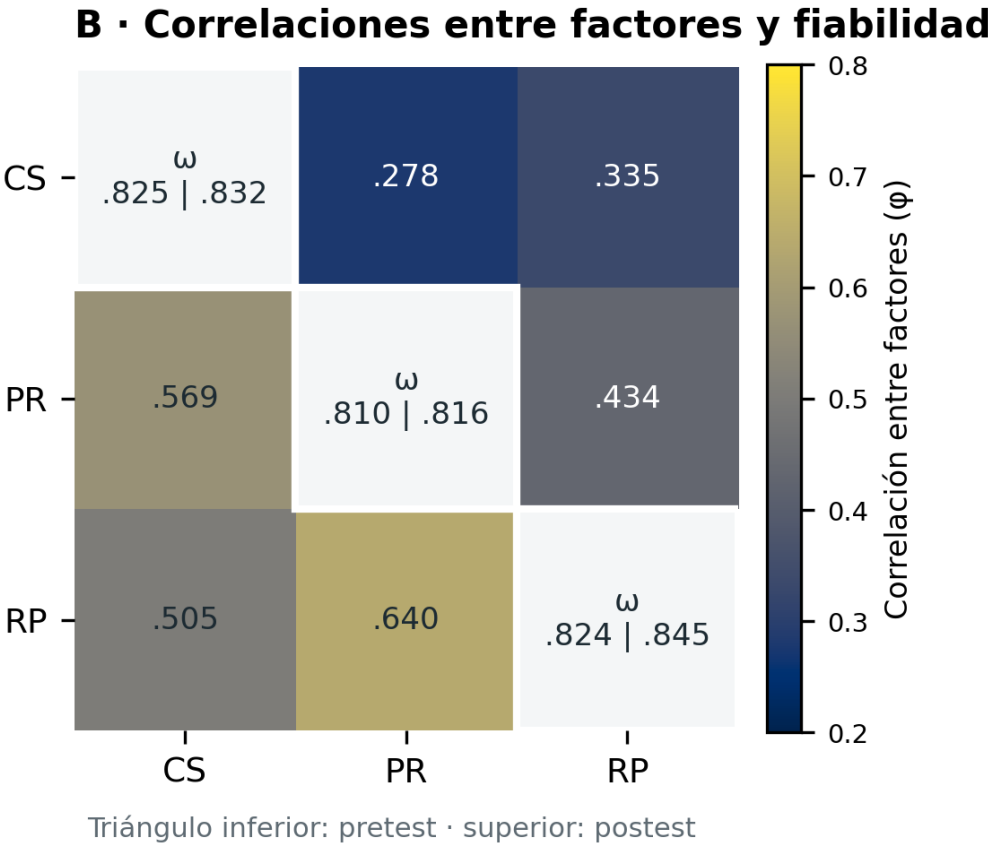
La Tabla 9-5 presenta los descriptivos por condición y medición. Las dos condiciones aumentaron su puntaje entre el pretest y el postest, con un incremento del total mayor en la condición experimental. La asimetría y la curtosis fueron inferiores a 1 en valor absoluto en todas las variables, y las correlaciones pretest-postest oscilaron entre .445 y .719.

Figura 9-1

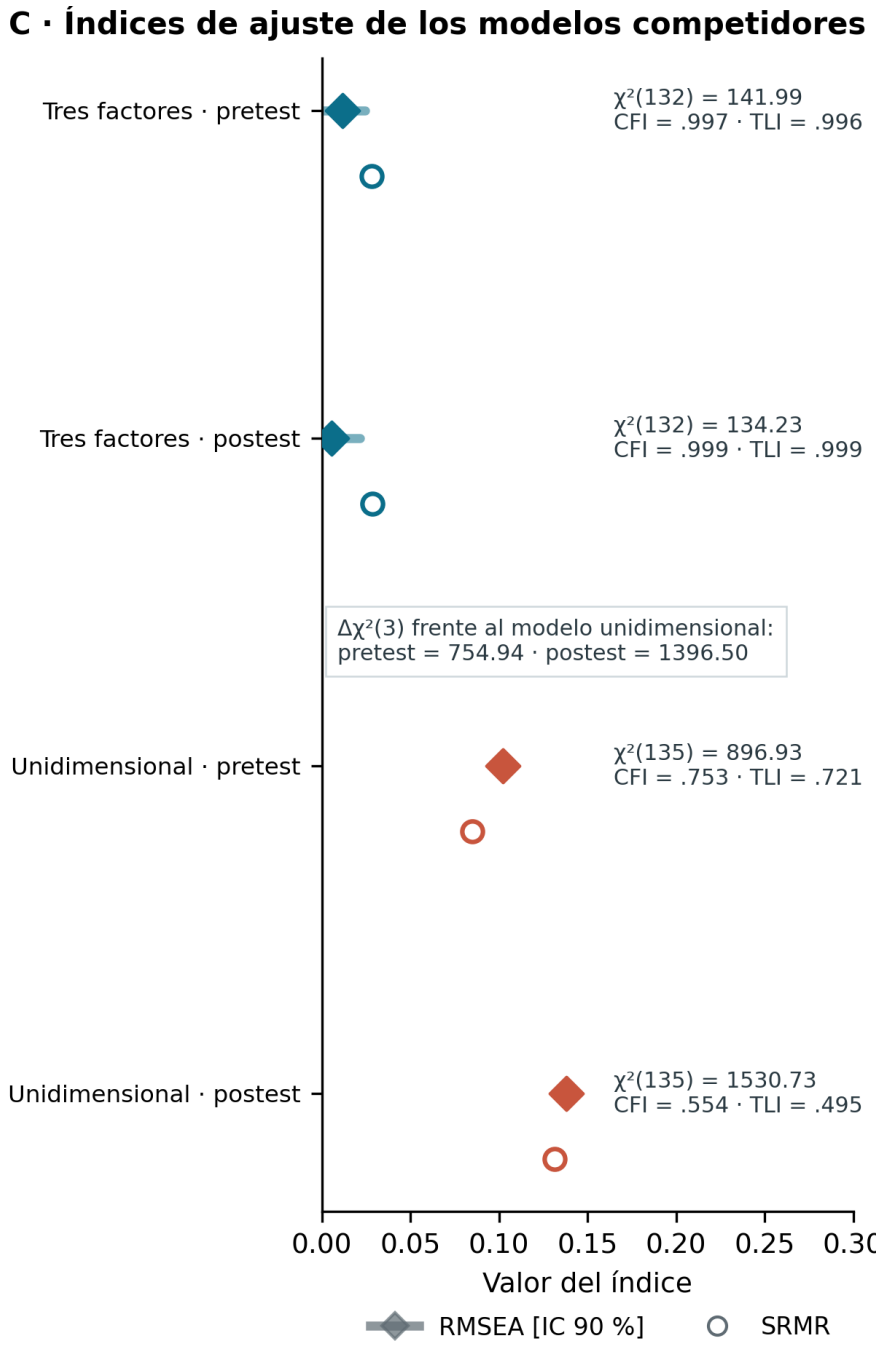
Estructura factorial de la prueba de razonamiento lógico-matemático en el pretest y el postest



Nota. N = 543; análisis factorial confirmatorio de tres factores correlacionados estimado por máxima verosimilitud. Panel A: cargas estandarizadas de los 18 ítems en el pretest (círculo vacío) y en el postest (círculo lleno), entre .579 y .745, con el ω de McDonald de cada dimensión (pretest / postest); la varianza media extraída de las dimensiones queda entre .418 y .477, por debajo de la mitad de la varianza de sus ítems



Nota. N = 543; análisis factorial confirmatorio de tres factores correlacionados estimado por máxima verosimilitud. Panel B: correlaciones entre factores en el pretest (triángulo inferior) y en el postest (triángulo superior), entre .278 y .640, y ω en la diagonal (.810 a .845).



Nota. N = 543; análisis factorial confirmatorio de tres factores correlacionados estimado por máxima verosimilitud. Panel C: RMSEA con su intervalo de confianza del 90 % (rombo y segmento) y SRMR (círculo) de los modelos de tres factores (pretest: RMSEA = .012, SRMR = .028; postest: RMSEA = .006, SRMR = .029) y unidimensional (pretest: RMSEA = .102, SRMR = .085; postest: RMSEA = .138, SRMR = .131), con χ^2, CFI y TLI de cada modelo; diferencia de χ^2 frente al modelo unidimensional: $\Delta\chi^2(3) = 754.94$ en el pretest y 1396.50 en el postest ($p < .001$).

Tabla 9-5

Estadísticos descriptivos del razonamiento lógico-matemático por condición y medición

Variable	Condición	pretest		postest		r pre–post
		M (DE)	Curtosis	M (DE)	Curtosis	
Clasificación y seriación	Experimental	8.26 (4.47)	–0.01 / –0.75	10.56 (4.39)	–0.19 / –0.78	.545
	Comparación	8.88 (4.13)	0.02 / –0.62	9.66 (4.36)	–0.22 / –0.71	.542
Patrones y regularidades	Experimental	9.15 (4.33)	–0.20 / –0.76	11.27 (4.20)	–0.37 / –0.73	.504
	Comparación	9.39 (4.34)	0.02 / –0.82	9.97 (4.29)	–0.04 / –0.86	.524
Resolución de problemas	Experimental	8.13 (4.28)	0.07 / –0.71	10.37 (4.40)	–0.22 / –0.87	.544
	Comparación	8.41 (4.38)	0.11 / –0.82	8.87 (4.67)	–0.01 / –0.97	.445
Puntaje total	Experimental	25.54 (10.51)	–0.14 / –0.39	32.19 (9.42)	–0.26 / –0.50	.719
	Comparación	26.68 (10.33)	–0.02 / –0.53	28.50 (9.51)	–0.00 / –0.53	.677

Nota: n experimental = 272; n comparación = 271. Dimensiones de 0 a 18 puntos y puntaje total de 0 a 54. Curt. = curtosis en exceso.

9.4.4 Supuestos del análisis de covarianza

La interacción condición × pretest no fue significativa en ninguna variable ($p \geq .278$; Tabla 9-6). La prueba de Shapiro-Wilk rechazó la normalidad de los residuos en patrones y regularidades ($p = .003$) y en resolución de problemas ($p = .011$), aunque la asimetría y la curtosis residuales fueron inferiores a 0.6 en valor absoluto en todas las variables. La prueba de Levene solo señaló diferencia de varianzas en resolución de problemas ($p = .042$). La correlación intraclase residual entre aulas fue de .064 a .079 en las dimensiones y de .177 en el puntaje total.

Tabla 9-6

Verificación de los supuestos del análisis de covarianza por variable dependiente

Variable	Shapiro-Wilk W (p)	Asim. / Curt. residuos	Levene F (p)	Pendientes F(1, 539) (p)	r pre–post	CCI residual del aula
Clasificación y seriación	.995 (.074)	−0.12 / −0.43	0.06 (.799)	0.25 (.616)	.532	.064
Patrones y regularidades	.991 (.003)	−0.19 / −0.46	0.15 (.699)	0.17 (.683)	.504	.078
Resolución de problemas	.993 (.011)	−0.02 / −0.58	4.14 (.042)	1.18 (.278)	.480	.079
Puntaje total	.996 (.162)	−0.01 / −0.39	1.38 (.241)	0.14 (.711)	.673	.177

Nota: Residuos del modelo postest ~ condición + pretest. Levene con centrado en la mediana. Pendientes = interacción condición × pretest (homogeneidad de las pendientes de regresión). CCI = correlación intraclase de los residuos entre aulas, estimada con un modelo de intercepto aleatorio por aula.

9.4.5 Efecto de la intervención

El análisis multivariado mostró un efecto conjunto sobre las tres dimensiones, controlando los tres pretest (traza de Pillai = .097, $F(3, 536) = 19.28$, $p < .001$). En el puntaje total, la condición experimental superó a la de comparación en 4.42 puntos ajustados (EE robusto = 1.26, $t(22) = 3.51$, $p = .002$), por lo que H1 se sostuvo. El efecto fue significativo en las tres dimensiones tras el ajuste de Holm, de modo que H2a, H2b y H2c también se sostuvieron (Tabla 9-7).

Tabla 9-7

Análisis multivariado y univariados de covarianza del efecto de la intervención sobre el postest, controlando el pretest

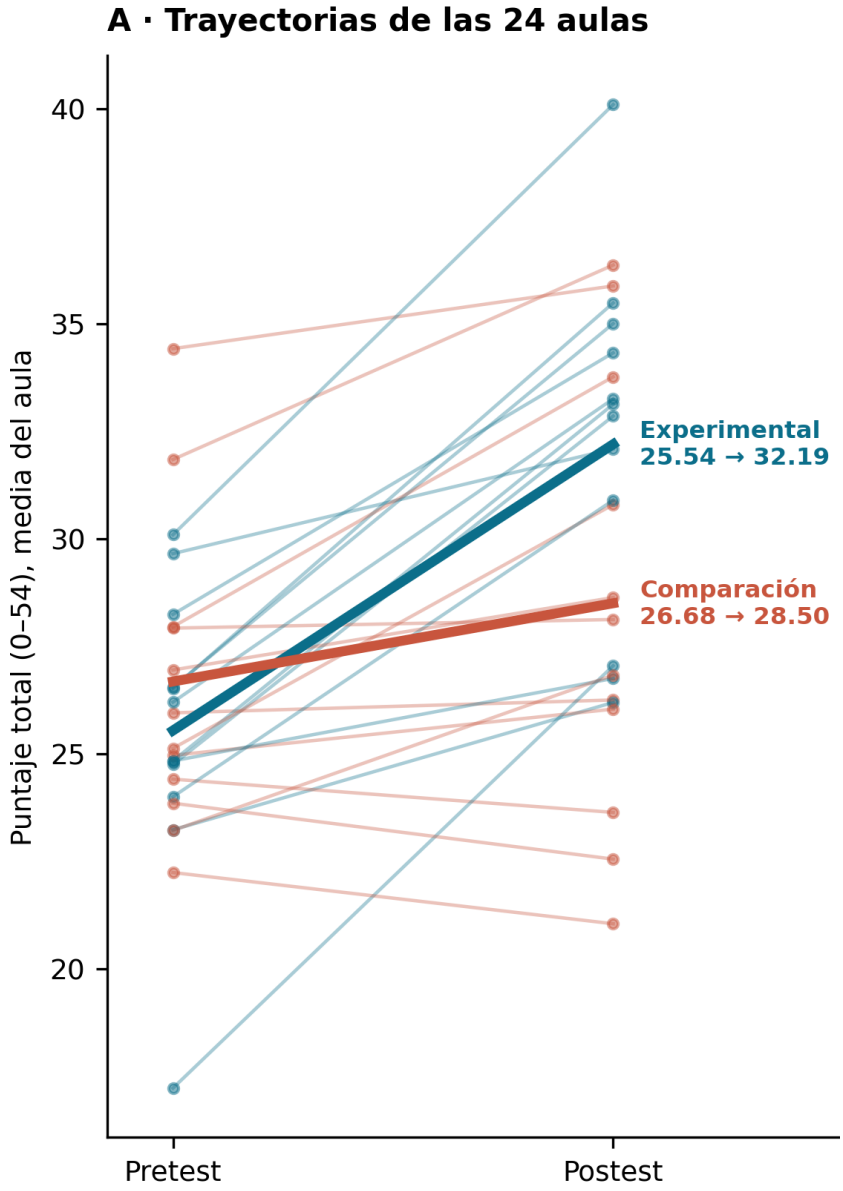
Variable dependiente	F	gl	p	η^2p [IC 95 %]	b (EE robusto)	t(22)	p	
							robusto	Holm
Multivariado (tres dimensiones): traza de Pillai = .097	19.28	3, 536	< .001	.097	—	—	—	—
Clasificación y seriación	15.33	1, 540	< .001	.028 [.007, .060]	1.24 (0.48)	2.58	.017	.020
Patrones y regularidades	20.72	1, 540	< .001	.037 [.012, .073]	1.42 (0.50)	2.83	.010	.020
Resolución de problemas	23.35	1, 540	< .001	.041 [.015, .079]	1.64 (0.54)	3.02	.006	.019
Puntaje total	57.39	1, 540	< .001	.096 [.054, .145]	4.42 (1.26)	3.51	.002	—

Nota: F y η^2p de los modelos de covarianza con el pretest de cada variable como covariable (en el análisis multivariado, los tres pretest). b = diferencia ajustada experimental – comparación; EE robusto por aula (CR1) e inferencia t con gl = J – 2 = 22 (J = 24 aulas). p Holm = ajuste de Holm sobre las tres dimensiones a partir del p robusto. El puntaje total es la hipótesis principal y no entra en el ajuste.

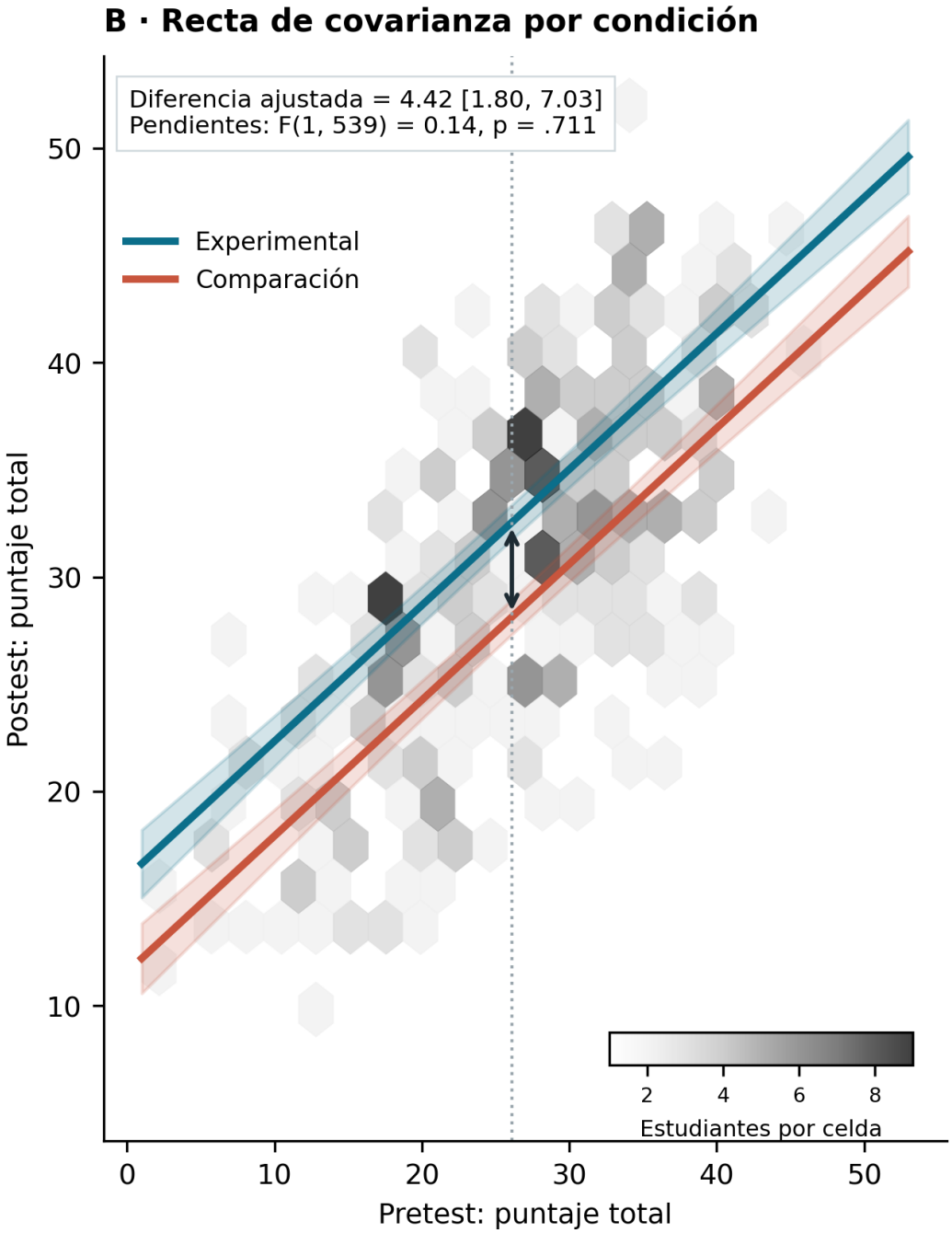
En términos de magnitud (Tabla 9-8 y Figura 9-2), el tamaño del efecto ajustado fue g = 0.47 en el puntaje total y osciló entre 0.28 y 0.36 en las dimensiones, con el valor mayor en resolución de problemas y el menor en clasificación y seriación; los intervalos de las tres dimensiones se solapan. La ganancia normalizada del total fue .234 en la condición experimental y .067 en la de comparación.

Figura 9-2

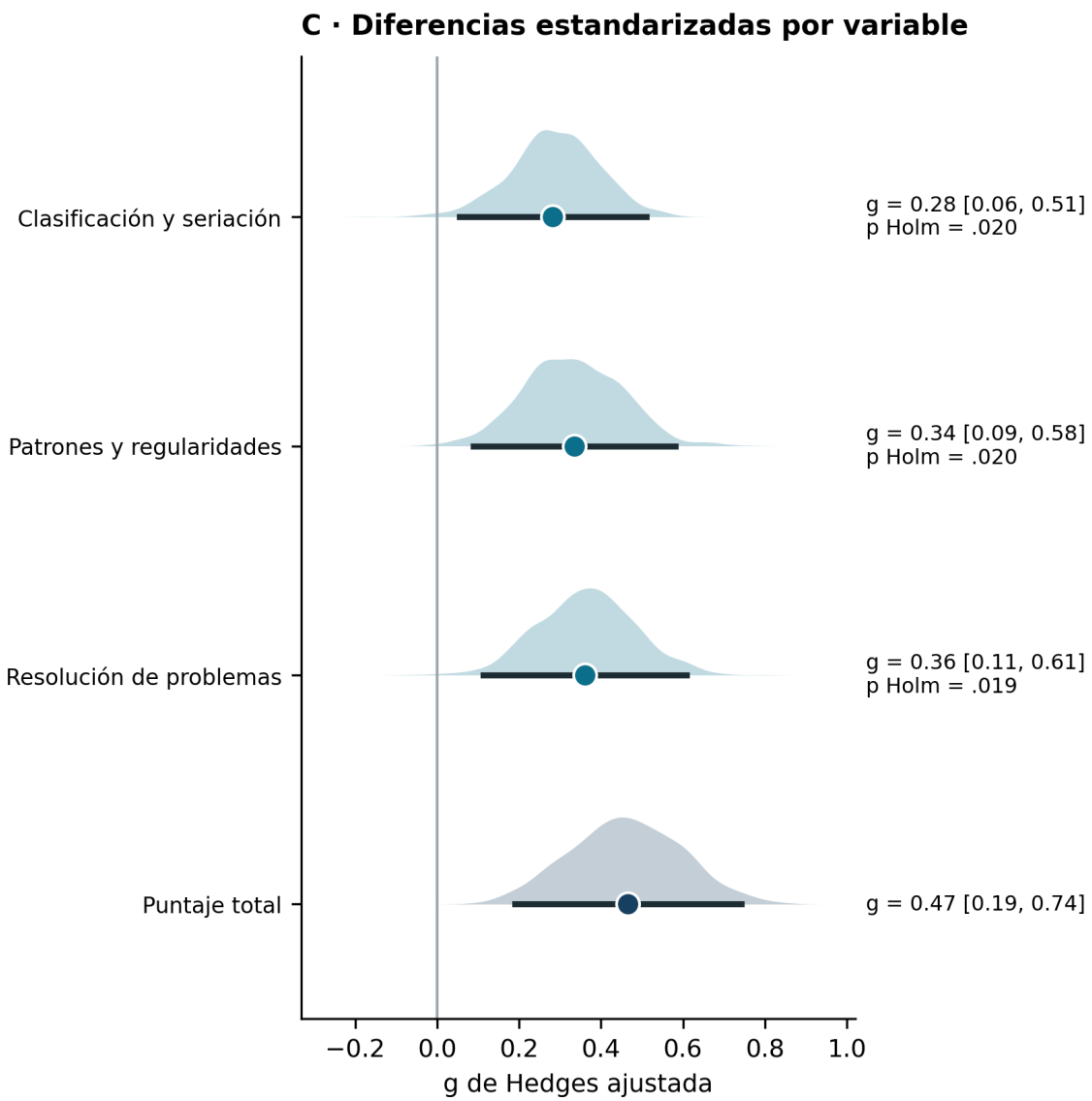
Efecto de la intervención con herramientas digitales sobre el razonamiento lógico-matemático



Nota. N = 543 (272 en la condición experimental y 271 en la de comparación) en 24 aulas; análisis de covarianza del postest controlando el pretest, con errores estándar robustos por aula e inferencia t(22). Panel A: media del puntaje total de cada aula en el pretest y el postest (líneas finas) y media de cada condición (líneas gruesas): experimental de 25.54 a 32.19 y comparación de 26.68 a 28.50.



Nota. N = 543 (272 en la condición experimental y 271 en la de comparación) en 24 aulas; análisis de covarianza del postest controlando el pretest, con errores estándar robustos por aula e inferencia t(22). Panel B: densidad de estudiantes por celda (hexágonos) y rectas del análisis de covarianza del puntaje total por condición con su banda de confianza del 95 %; la flecha marca la diferencia ajustada en la media del pretest, 4.42 puntos [IC 95 % 1.80, 7.03]; interacción condición × pretest F(1, 539) = 0.14, p = .711.



Nota. N = 543 (272 en la condición experimental y 271 en la de comparación) en 24 aulas; análisis de covarianza del postest controlando el pretest, con errores estándar robustos por aula e inferencia t(22). Panel C: g de Hedges ajustada con su intervalo de confianza del 95 % para cada dimensión y para el puntaje total (0.47 [0.19, 0.74]), sobre la distribución de 1 000 remuestreos de aulas dentro de cada condición; a la derecha, el valor de g y, en las dimensiones, el p ajustado por Holm.s

Tabla 9-8

Medias ajustadas por el pretest, diferencia ajustada, tamaño del efecto y ganancia normalizada por condición

Variable	M ajustada		Dif.	g de Hedges ajustada	Ganancia normalizada	
	experimental	comparación	[IC 95 %]		experimental	comparación
Clasificación y seriación	10.73	9.49	1.24 [0.24, 2.24]	0.28 [0.06, 0.51]	.236	.086
Patrones y regularidades	11.33	9.91	1.42 [0.38, 2.47]	0.34 [0.09, 0.58]	.240	.067
Resolución de problemas	10.44	8.80	1.64 [0.51, 2.76]	0.36 [0.11, 0.61]	.227	.048
Puntaje total	32.56	28.14	4.42 [1.80, 7.03]	0.47 [0.19, 0.74]	.234	.067

Nota: Medias ajustadas al promedio muestral del pretest. Intervalos con EE robusto por aula y t(22). g ajustada = diferencia ajustada / DE combinada del postest, con corrección de Hedges. Ganancia normalizada = (M postest − M pretest)/(máximo − M pretest)

9.4.6 Robustez

El efecto se mantuvo con todos los estimadores alternativos (Tabla 9-9 y Figura 9-3, paneles A y B). Para el puntaje total, el modelo multinivel con efectos fijos de institución estimó b = 4.52, la ANCOVA con efectos fijos de institución b = 4.39, el bootstrap salvaje por aula dio p = .002 y el bootstrap de aulas situó el efecto entre 1.87 y 6.83 puntos. El análisis por rangos de Quade fue significativo en todas las variables y, al excluir cada institución, el efecto sobre el total varió entre 3.62 y 4.95 puntos. En las tres dimensiones, todos los estimadores conservaron el signo y la significación del análisis principal.

Tabla 9-9

Análisis de sensibilidad del efecto de la intervención con estimadores e inferencias alternativos

Variable	Multinivel: b (EE), p	CCI del aula	Efectos fijos de institución: b (EE), p	Bootstrap aulas		Quade: F(1, 541), p	Exclusión de una institución: rango de b
				salvaje: p	IC 95 % de b		
Clasificación y seriación	1.24 (0.45), .006	.047	1.23 (0.38), .005	.021	[0.33, 2.17]	12.23, < .001	0.81 a 1.55
Patrones y regularidades	1.46 (0.49), .003	.069	1.42 (0.41), .003	.011	[0.42, 2.42]	19.46, < .001	1.16 a 1.72
Resolución de problemas	1.69 (0.52), .001	.064	1.63 (0.44), .002	.008	[0.61, 2.65]	20.55, < .001	1.29 a 1.94
Puntaje total	4.52 (1.17), < .001	.143	4.39 (0.97), < .001	.002	[1.87, 6.83]	53.50, < .001	3.62 a 4.95

Nota: Multinivel = modelo de intercepto aleatorio por aula con efectos fijos de institución (REML). Efectos fijos de institución = ANCOVA con cinco indicadores de institución, EE robusto por aula y t(17). Bootstrap salvaje con hipótesis nula impuesta y pesos de seis puntos de Webb (9999 réplicas). Bootstrap de aulas = remuestreo de aulas dentro de cada condición (2000 réplicas, intervalo percentil). Quade = análisis de covarianza por rangos. Exclusión de una institución = estimación de b retirando cada una de las seis instituciones.

9.4.7 Niveles de logro

La Tabla 9-10 y el panel C de la Figura 9-3 muestran el desplazamiento de los estudiantes entre niveles de logro. En la condición experimental, la proporción en el nivel satisfactorio pasó de 11.8 % a 30.5 % y la proporción en inicio bajó de 33.1 % a 15.4 % (Stuart-Maxwell $\chi^2(2) = 69.61$, $p < .001$); en la de comparación, el nivel satisfactorio pasó de 14.8 % a 18.8 % ($\chi^2(2) = 16.86$, $p < .001$). Las condiciones no diferían en la distribución de niveles en el pretest ($p = .488$), pero sí en el postest ($\chi^2(2) = 12.17$, $p = .002$, $V = .106$).

Tabla 9-10

Distribución de los niveles de logro del puntaje total en el pretest y el postest por condición

Condición	Nivel en el pretest	En inicio (postest)	En proceso (postest)	Satisfactorio (postest)	Total pretest
Experimental	En inicio	36	48	6	90
	En proceso	6	90	54	150
	Satisfactorio	0	9	23	32
	Total postest	42	147	83	272
Comparación	En inicio	47	44	2	93
	En proceso	16	97	25	138
	Satisfactorio	0	16	24	40
	Total postest	63	157	51	271

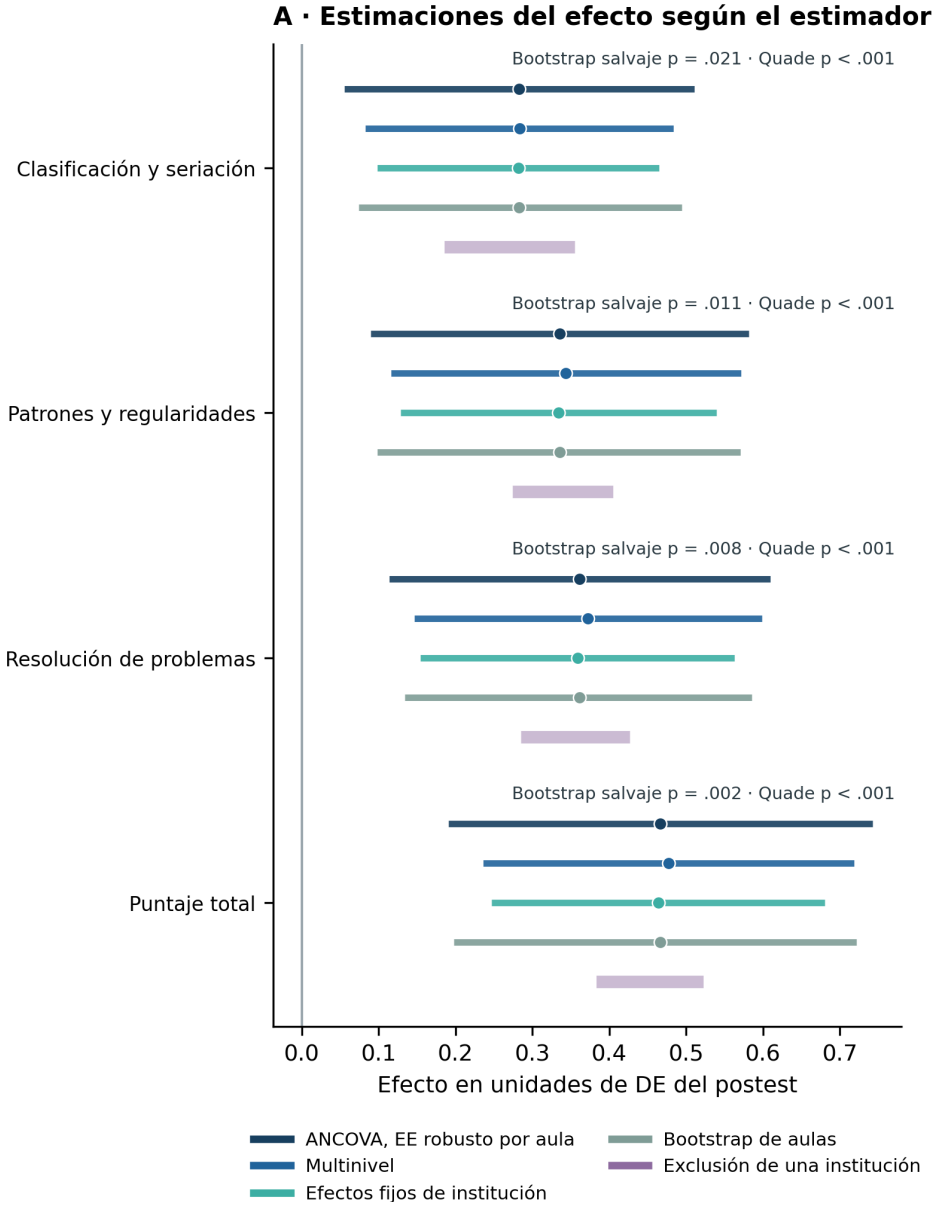
Nota: Niveles del puntaje total (0–54): en inicio ≤ 21; en proceso 22–37; satisfactorio ≥ 38. Prueba de homogeneidad marginal de Stuart-Maxwell (pretest frente a postest): experimental $\chi^2(2) = 69.61$, $p < .001$; comparación $\chi^2(2) = 16.86$, $p < .001$. Distribución en el postest entre condiciones: $\chi^2(2) = 12.17$, $p = .002$, V de Cramér = .106 (pretest: $\chi^2(2) = 1.44$, $p = .488$).

9.5 DISCUSIÓN

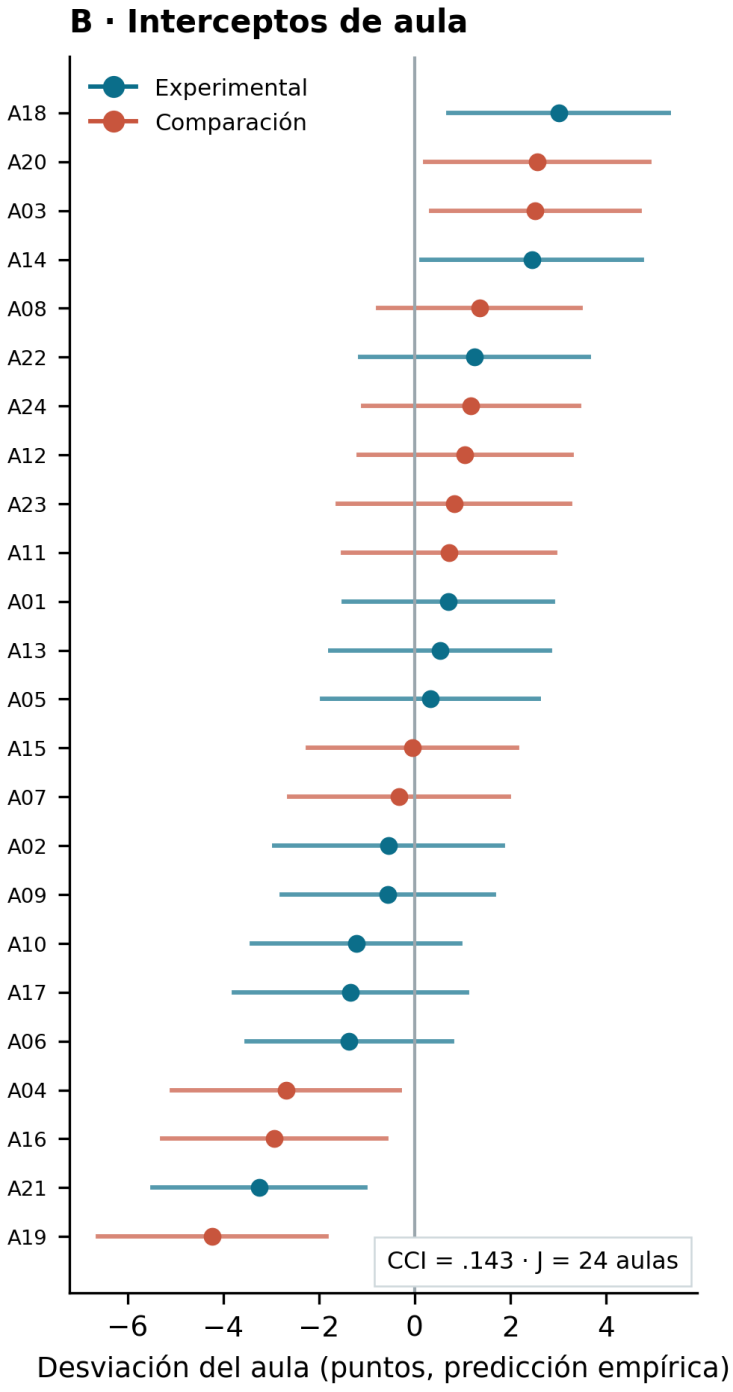
El hallazgo más desconcertante no es que la intervención mejorara el razonamiento, sino que lo hiciera con igual tiempo lectivo en ambas condiciones y en un país donde la multiplicación de computadoras en 318 escuelas primarias rurales no mostró efectos sobre las pruebas de matemática y solo ofreció indicios inconclusos sobre las habilidades cognitivas generales (Cristia et al., 2017). La diferencia ajustada de $g = 0.47$ en el puntaje total se sitúa cerca del efecto medio de la tecnología en la matemática de primaria (Akçay et al., 2021), por debajo del estimado para secundaria (Hillmayr et al., 2020) y claramente por encima del de las herramientas de seguimiento docente (Faber et al., 2023); frente a las referencias empíricas de intervenciones evaluadas en el terreno, cuyos efectos suelen ser menores que los umbrales convencionales, una magnitud como esta se considera grande (Kraft, 2020).

Figura 9-3

Robustez del efecto, variación entre aulas y transiciones entre niveles de logro

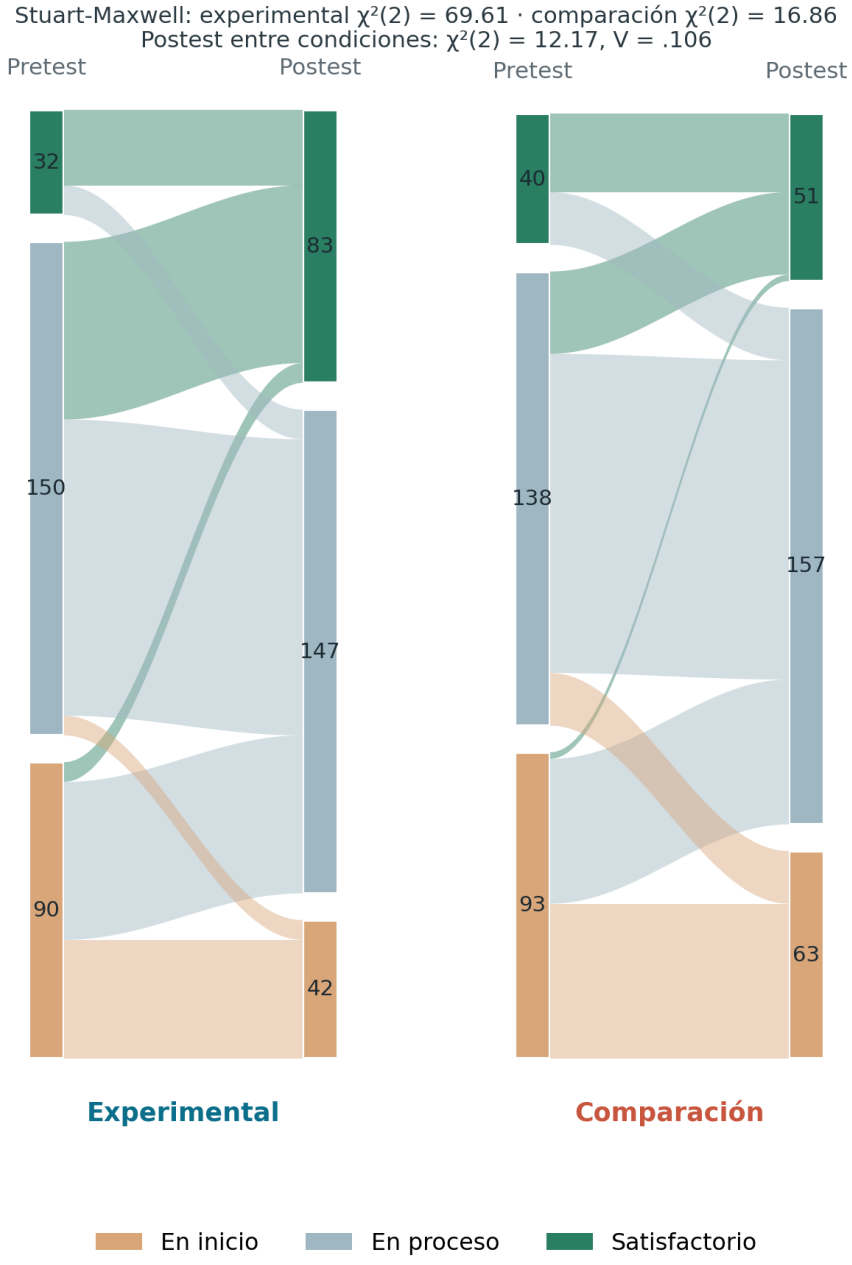


Nota. N = 543 en 24 aulas. Panel A: efecto de la intervención en unidades de desviación estándar del postest para cada variable según cinco estimaciones: análisis de covarianza con errores robustos por aula, modelo multinivel de intercepto aleatorio (intervalo de Wald del 95 %), análisis de covarianza con efectos fijos de institución (t(17)), bootstrap de aulas (2000 réplicas, intervalo percentil) y rango de estimaciones al excluir cada institución; sobre cada bloque, el p del bootstrap salvaje por aula (9999 réplicas, pesos de Webb) y del análisis de covarianza por rangos de Quade.



Nota. N = 543 en 24 aulas. Panel B: desviación de cada aula respecto de la predicción del modelo multinivel del puntaje total (con efectos fijos de institución), con su intervalo del 95 %; correlación intraclase = .143.

C · Transiciones entre niveles de logro



Nota. N = 543 en 24 aulas. Panel C: flujos de estudiantes entre los niveles de logro del puntaje total (en inicio ≤ 21; en proceso 22–37; satisfactorio ≥ 38) del pretest al postest en cada condición, con el número de estudiantes por nivel; Stuart-Maxwell experimental $\chi^2(2) = 69.61$ y comparación $\chi^2(2) = 16.86$; distribución del postest entre condiciones $\chi^2(2) = 12.17$, V de Cramér = .106..

Estos contrastes sugieren que el beneficio no procede de la disponibilidad de dispositivos, sino de su integración en tareas dirigidas a procesos de razonamiento concretos y de la formación y el acompañamiento de las docentes, aunque el diseño no permite aislar la contribución de cada ingrediente. El desplazamiento hacia el nivel satisfactorio, que más que se duplicó en la condición experimental frente a un aumento de cuatro puntos porcentuales en la de comparación, apunta en la misma dirección y es consistente con que el efecto no se limite a un corrimiento de la media. Cabe señalar, además, que las diferencias iniciales favorecían ligeramente a la condición de comparación, de manera que el ajuste por el pretest actuó en contra de la intervención y no a su favor.

Desde una perspectiva crítica, esa magnitud no debería leerse como una promesa generalizable al conjunto de la región. La evidencia latinoamericana con datos de PISA muestra que el acceso y el uso de la tecnología se relacionan de forma desigual con los resultados: la disponibilidad y el uso de las TIC se han puesto en relación con la eficiencia de 2757 escuelas de diez países (Agasisti et al., 2023); los recursos digitales del hogar amplifican el gradiente socioeconómico en matemática en 13 sistemas de la región (Torres et al., 2026); y el diagnóstico de equidad de PISA 2022 sitúa los correlatos digitales junto a los socioeconómicos y escolares como parte de una misma estructura de desigualdad (Guerra et al., 2026). La eficacia de las políticas de aprendizaje remoto durante la pandemia tampoco fue homogénea (Delprato, 2026), y en escuelas rurales chilenas se documentó que, aunque el programa proveyó computadoras, persistían la señal deficiente, la formación docente insuficiente y las destrezas digitales limitadas (Pavez y Farías, 2026). Nuestras aulas, públicas, con un aula de innovación operativa y reclutadas por conveniencia, representan condiciones de implementación más favorables que las descritas en esos trabajos, de modo que la transferencia a contextos con menos infraestructura constituye una hipótesis por contrastar y no una inferencia. Tampoco se estimaron efectos diferenciales por sexo, pese a que las brechas de género en matemática en América Latina y el Caribe son amplias y desfavorecen a las niñas (Ortega et al., 2026); la composición

por sexo fue similar entre condiciones, pero esa similitud no informa sobre si el beneficio fue equitativo. Estos límites cobran peso porque un efecto tecnológico que depende de la infraestructura y de la formación docente podría ampliar, y no reducir, las brechas existentes.

La coincidencia más directa es la de un cuasiexperimento en secundaria en el que la programación con Scratch y Python produjo mayores ganancias en razonamiento lógico y numérico en el grupo experimental (Pajares et al., 2026), resultado compatible con el nuestro aunque con otro nivel educativo, otra herramienta y otro instrumento.

La mayor magnitud en resolución de problemas y en patrones que en clasificación y seriación admite una lectura apoyada en las tres vías descritas en el marco teórico. La práctica adaptativa y los juegos de lógica, concentrados en el último bloque, ofrecen más ocasiones de retroalimentación sobre secuencias de pasos que sobre operaciones de clasificación, que en quinto grado probablemente están más consolidadas; la relación entre el reconocimiento de patrones y la resolución de problemas en primaria, aunque descrita en trabajos exploratorios (Rizos y Gkrekas, 2024), apunta en la misma dirección. Rasgos de diseño como el número de intentos que permite un juego de práctica matemática también modulan el aprendizaje y la motivación (Cubillos et al., 2023), y podrían explicar parte de la variación entre componentes si las tareas difirieron en ese aspecto, algo que no se midió. Con todo, los intervalos de las tres dimensiones se solapan y el orden de los bloques coincide con el de las dimensiones, por lo que el diseño no separa el contenido del componente del momento en que se trabajó; esa distinción exige diseños con orden contrabalanceado.

Para la práctica, los resultados sugieren que las herramientas digitales mejoran el razonamiento cuando cada una se alinea con un proceso concreto —manipulación con retroalimentación para clasificar y seriar, geometría dinámica para generalizar regularidades y práctica adaptativa para resolver problemas— y cuando las docentes cuentan con formación y acompañamiento. La selección de aplicaciones debería atender a criterios

de aprendizaje activo, implicación, significatividad e interacción social más que a su popularidad, dado que la calidad educativa de muchas aplicaciones comerciales es variable (Meyer et al., 2021). La literatura reciente converge en que los rasgos de diseño de la herramienta, y no solo su presencia, condicionan el beneficio: las síntesis sobre juegos digitales en la matemática de primaria (Dan et al., 2024), sobre la tecnología como recurso para reducir la ansiedad matemática (Ersozlu, 2024) y sobre la educación matemática asistida por robots (Samuel et al., 2025) amplían el repertorio de herramientas disponibles, mientras que un experimento con asignación aleatoria sobre los mecanismos de competición se pregunta si estos solo elevan el desempeño dentro del juego o favorecen un dominio genuino del conocimiento (Chen et al., 2026). En el campo del pensamiento computacional, la comparación entre programación con y sin computadora (Sun et al., 2024) y el análisis de las características del estudiante y de la escuela (Küçükaydın y Çite, 2024) indican que la modalidad y el contexto modulan lo que la herramienta aporta. La integración de herramientas digitales en modelos como el aula invertida de primaria descansa, asimismo, en las experiencias y percepciones de docentes y estudiantes (Loizou, 2022), un aspecto que este estudio no recogió y que podría complementar la dosis de uso registrada en cada aula.

Para la investigación, los datos muestran que la correlación intraclase residual de las aulas no es despreciable aun después de controlar el pretest, y que la inferencia con pocos conglomerados exige más aulas de las que sugiere el cálculo analítico corriente: la simulación con el análisis declarado situó el umbral de potencia en 12 aulas por condición, frente a las 8 del cálculo analítico. Planificar los cuasiexperimentos educativos con el aula como unidad y verificar la potencia por simulación con el análisis que se empleará es una práctica que el campo haría bien en adoptar, puesto que ignorar la estructura por aulas puede subestimar los errores estándar y conducir a conclusiones más seguras de lo que los datos autorizan.

Las limitaciones acotan la interpretación causal. La asignación no fue aleatoria; aunque se realizó dentro de cada institución y las diferencias iniciales fueron pequeñas, no pueden descartarse diferencias no medidas

entre las docentes de cada condición, que están confundidas con la condición en el nivel del aula. Los valores de planificación resultaron optimistas: la correlación pretest-postest observada en el puntaje total fue menor que la prevista (.60) y la correlación intraclase residual de las dimensiones (.064 a .079) superó la supuesta (.05), por lo que la precisión alcanzada es menor que la planificada y los intervalos de las dimensiones son amplios. El instrumento se elaboró para el estudio y su estructura se verificó en la misma muestra, y los ítems de crédito parcial se trataron como continuos. La intervención duró doce semanas, se evaluó en un solo grado y sin medición de seguimiento, de modo que se desconoce la persistencia del efecto. Las líneas futuras incluyen ensayos con asignación aleatoria de aulas o instituciones, medición diferida, invarianza de la medida entre condiciones, análisis de la dosis de uso y del papel de la ansiedad matemática, que aparece ya en el primer año de la escolarización primaria y se asocia negativamente con el rendimiento (Szczygieł y Pieronkiewicz, 2022).

9.6 CONCLUSIONES

Una intervención de doce semanas con herramientas digitales alineadas con los componentes del razonamiento mejoró el razonamiento lógico-matemático de estudiantes de quinto grado de primaria, con un efecto de g = 0.47 en el puntaje total y efectos en la clasificación y la seriación, los patrones y regularidades y la resolución de problemas, robustos a la estructura por aulas y a los estimadores alternativos. Las hipótesis H1, H2a, H2b y H2c se sostuvieron, y el desplazamiento de los estudiantes hacia el nivel satisfactorio fue mayor en la condición experimental que en la de comparación.

El resultado sugiere que el aporte de la tecnología al aprendizaje de la matemática depende de su integración didáctica en tareas de razonamiento y de la preparación de las docentes, y no del acceso a dispositivos. Para los sistemas educativos, ello desplaza la prioridad desde la dotación de equipos hacia el diseño de secuencias que alineen cada herramienta con

un proceso de razonamiento, la formación y el acompañamiento docente, y la selección de aplicaciones por su calidad pedagógica.

Para la investigación, el estudio ofrece una estimación con inferencia por conglomerados y con potencia verificada por simulación, que la evidencia cuasiexperimental del campo rara vez proporciona, y un instrumento cuya estructura factorial se sostuvo en las dos mediciones. Su alcance está acotado por la asignación no aleatoria de las aulas, por la ausencia de seguimiento y por el contexto de implementación favorable de las instituciones participantes, por lo que la generalización a escuelas con menos infraestructura, la persistencia del efecto y su equidad entre grupos de estudiantes quedan como tareas pendientes para ensayos con asignación aleatoria de aulas y medición diferida.

9.7 REFERENCIAS

Agasisti, T., Antequera, G., & Delprato, M. (2023). Technological resources, ICT use and schools efficiency in Latin America – Insights from OECD PISA 2018. *International Journal of Educational Development, 99*, 102757. https://doi.org/10.1016/j.ijedudev.2023.102757

Ahmad, S., & Siller, H. (2024). Investigating the effect of manipulatives on mathematics achievement: The role of concrete and virtual manipulatives for diverse achievement level groups. *Journal on Mathematics Education, 15*(3), 979-1002. https://doi.org/10.22342/jme.v15i3.pp979-1002

Akçay, A., Karahan, E., & Bozan, M. (2021). The Effect of Using Technology in Primary School Math Teaching on Students' Academic Achievement: A Meta-Analysis Study. *FIRE: Forum for International Research in Education, 7*(2), 1-21. https://doi.org/10.32865/fire202172231

Ancheta, A., Pulido, C., & Carvajal, V. (2021). Gender Digital Divide and Education in Latin America: A Literature Review. *Education Sciences, 11*(12), 804. https://doi.org/10.3390/educsci11120804

Behnamnia, N., Kamsin, A., & Hayati, S. (2024). Impact of Digital Game-Based Learning on STEM education in primary schools: a meta-analysis of learning approaches. *Innoeduca. International Journal of Technology and Educational Innovation, 10*(2), 113-140. https://doi.org/10.24310/ijtei.102.2024.19694

Cai, Z., Zhang, X., Liu, C., & Zhan, J. (2025). Effects of Digital Game-Based Learning on Student's Problem-Solving Ability: A Three-Level Meta-Analysis. *Journal of Computer Assisted Learning, 41*(2). https://doi.org/10.1111/jcal.70002

Charalambous, Y., & Charalambous, C. (2025). Examining the effects of an intervention on mathematical modeling in problem solving at upper elementary grades: a cluster randomized trial study. *Mathematical Thinking and Learning, 27*(2), 202-220. https://doi.org/10.1080/10986065.2023.2270088

Chen, Z., Huang, Q., Zheng, B., Mu, M., Zhao, Y., Li, Y., & Zhou, D. (2026). The impact of competition mechanisms on primary school students' digital game-based mathematics learning. *Acta Psychologica, 267*, 107175. https://doi.org/10.1016/j.actpsy.2026.107175

Coskun, K., & Kara, C. (2022). The impact of the COVID-19 pandemic on primary school students' mathematical reasoning skills: a mediation analysis. *London Review of Education, 20*(1). https://doi.org/10.14324/lre.20.1.19

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2017). Technology and Child Development: Evidence from the One Laptop per Child Program. *American Economic Journal: Applied Economics, 9*(3), 295-320. https://doi.org/10.1257/app.20150385

Cubillos, C., Roncagliolo, S., & Cabrera, D. (2023). Learning and Motivation When Using Multiple-Try in a Digital Game for Primary Students in Chile. *Education Sciences, 13*(11), 1119. https://doi.org/10.3390/educsci13111119

Cueto, S., Beuermann, D., Cristia, J., Malamud, O., & Pardo, F. (2025). Laptops in the long run: Evidence from the one laptop per child program in rural Peru. *Journal of Public Economics, 252*, 105538. https://doi.org/10.1016/j.jpubeco.2025.105538

Cui, Z., & Ng, O. (2021). The Interplay Between Mathematical and Computational Thinking in Primary School Students' Mathematical Problem-Solving Within a Programming Environment. *Journal of Educational Computing Research, 59*(5), 988-1012. https://doi.org/10.1177/0735633120979930

Dan, N., Trung, L., Nga, N., & Dung, T. (2024). Digital game-based learning in mathematics education at primary school level: A systematic literature review. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(4), em2423. https://doi.org/10.29333/ejmste/14377

Debrenti, E. (2024). Game-Based Learning experiences in primary mathematics education. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1331312

Delprato, M. (2026). Effects of COVID-19 education policies on students' learning proficiency in Latin America—Insights from PISA 2022. *International Journal of Educational Development, 120,* 103483. https://doi.org/10.1016/j.ijedudev.2025.103483

Ersozlu, Z. (2024). The role of technology in reducing mathematics anxiety in primary school students. *Contemporary Educational Technology, 16*(3), ep517. https://doi.org/10.30935/cedtech/14717

Faber, J., Feskens, R., & Visscher, A. (2023). A best-evidence meta-analysis of the effects of digital monitoring tools for teachers on student achievement. *School Effectiveness and School Improvement, 34*(2), 169-188. https://doi.org/10.1080/09243453.2022.2142247

Guerra, J., Roblez, G., Vargas, V., Meneses, M., & Basantes, A. (2026). Digital equity and socioeconomic gradients in PISA 2022 mathematics: an integrated equity-diagnostic framework for Latin America and the Caribbean. *Frontiers in Education, 11.* https://doi.org/10.3389/feduc.2026.1810210

Hedges, L., & Hedberg, E. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87. https://doi.org/10.3102/0162373707299706

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S., & Reiss, K. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153,* 103897. https://doi.org/10.1016/j.compedu.2020.103897

Huang, F., & Li, X. (2021). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods, 54*(3), 1181-1199. https://doi.org/10.3758/s13428-021-01627-0

James, E., Jay, T., & Sella, F. (2024). No evidence that playing a linear number board game improves numerical skills beyond teaching as usual: A randomized controlled trial in 4- to 5-year-old primary school children. *Journal of Educational Psychology, 116*(3), 411-425. https://doi.org/10.1037/edu0000857

Kraft, M. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher, 49*(4), 241-253. https://doi.org/10.3102/0013189x20912798

Küçükaydın, M., & Çite, H. (2024). Computational thinking in primary school: effects of student and school characteristics. *Education and Information Technologies, 29*(5), 5631-5649. https://doi.org/10.1007/s10639-023-12052-5

Lee, Y., & Pustejovsky, J. (2024). Comparing random effects models, ordinary least squares, or fixed effects with cluster robust standard errors for cross-classified data. *Psychological Methods, 29*(6), 1084-1099. https://doi.org/10.1037/met0000538

Loizou, M. (2022). Digital Tools and the Flipped Classroom Approach in Primary Education. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.793450

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling.

Psychological Methods, 1(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Mamolo, L., & Sugano, S. (2024). Digital interactive app and students' mathematics self-efficacy, anxiety, and achievement in the "new normal". *E-Learning and Digital Media, 21*(5), 427-443. https://doi.org/10.1177/20427530231167646

Meyer, M., Zosh, J., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R., Hirsh, K., & Radesky, J. (2021). How educational are "educational" apps for young children? App store content analysis using the Four Pillars of Learning framework. *Journal of Children and Media, 15*(4), 526-548. https://doi.org/10.1080/17482798.2021.1882516

Nordby, S., Bjerke, A., & Mifsud, L. (2022). Computational Thinking in the Primary Mathematics Classroom: a Systematic Review. *Digital Experiences in Mathematics Education, 8*(1), 27-49. https://doi.org/10.1007/s40751-022-00102-5

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology, 25*(1), 147-166. https://doi.org/10.1348/026151006x153127

Ortega, L., Montero, M., Romero, Á., Canals, C., & Mizala, A. (2026). Pathways to gender equity in Latin America and the Caribbean: A comparative analysis of trends in mathematics gender gaps using PISA data. *Studies in Educational Evaluation, 88,* 101579. https://doi.org/10.1016/j.stueduc.2026.101579

Ouyang, Y., Hemming, K., Li, F., & Taljaard, M. (2023). Estimating intra-cluster correlation coefficients for planning longitudinal cluster randomized trials: a tutorial. *International Journal of Epidemiology*, *52*(5), 1634-1647. https://doi.org/10.1093/ije/dyad062

Öz, T., & Işık, A. (2024). Exploring mathematical reasoning skills of middle school students. *Thinking Skills and Creativity*, *53*, 101612. https://doi.org/10.1016/j.tsc.2024.101612

Özdemir, O. (2025). Kahoot! Game-based digital learning platform: A comprehensive meta-analysis. *Journal of Computer Assisted Learning*, *41*(1). https://doi.org/10.1111/jcal.13084

Pajares, R., Martín, L., & Carbonero, M. (2026). The Impact of Computational Thinking on Logical-Mathematical Reasoning in High School Education: A Quasi-Experimental Study. *Education Sciences*, *16*(2), 345. https://doi.org/10.3390/educsci16020345

Pavez, I., & Farías, C. (2026). One Laptop per Child: a case study of digital inclusion in rural Chile. *Technology, Pedagogy and Education*, *35*(2), 341-360. https://doi.org/10.1080/1475939x.2025.2541958

Pratama, R. (2025). Effectiveness of research and development (R&D) of android-based educational games to improve students' mathematical reasoning: a meta-analysis approach. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00486-7

Rizos, I., & Gkrekas, N. (2024). Pattern recognition among primary school students: The relationship with mathematical problem-solving. *Contemporary Mathematics and Science Education*, *5*(2), ep24010. https://doi.org/10.30935/conmaths/14689

Samuel, T., Chen, H., & Yohannes, A. (2025). Robot-Assisted Mathematics Education: A Systematic Literature Review. *Journal of Computer Assisted Learning*, *41*(4). https://doi.org/10.1111/jcal.70095

Santaria, Y., & Mamani, J. (2026). Ansiedad matemática, autoeficacia y nivel socioeconómico como predictores del rendimiento matemático de estudiantes peruanos con datos PISA. *Pesquimat*, *29*(1), 63-85. https://doi.org/10.15381/pesquimat.v29.i1.33140

Schaaf, J., Rolfes, T., Nagy, G., & Heinze, A. (2026). The effect of the frequency of use of an intelligent tutoring system on learning gains in mathematics secondary education. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1738655

Scippo, S., Madiai, S., & Cuomo, S. (2025). Digital Tessellation for Geometry Learning in Primary School: A Quasi-Experimental Study. *Journal of Information Technology Education: Research*, *24*, 009. https://doi.org/10.28945/5473

Son, T. (2024). Intelligent Tutoring Systems in Mathematics Education: A Systematic Literature Review Using the Substitution, Augmentation, Modification, Redefinition Model. *Computers*, *13*(10), 270. https://doi.org/10.3390/computers13100270

Spitzer, M., & Moeller, K. (2024). Performance increases in mathematics within an intelligent tutoring system during COVID-19 related school closures: a large-scale longitudinal evaluation. *Computers and Education Open*, *6*, 100162. https://doi.org/10.1016/j.caeo.2024.100162

Sun, L., Liu, J., & Liu, Y. (2024). Comparative experiment of the effects

of unplugged and plugged-in programming on computational thinking in primary school students: A perspective of multiple influential factors. *Thinking Skills and Creativity*, *52*, 101542. https://doi.org/10.1016/j.tsc.2024.101542

Sun, L., Ruokamo, H., Siklander, P., Li, B., & Devlin, K. (2021). Primary school students' perceptions of scaffolding in digital game-based learning in mathematics. *Learning, Culture and Social Interaction*, *28*, 100457. https://doi.org/10.1016/j.lcsi.2020.100457

Szczygieł, M., & Pieronkiewicz, B. (2022). Exploring the nature of math anxiety in young children: Intensity, prevalence, reasons. *Mathematical Thinking and Learning*, *24*(3), 248-266. https://doi.org/10.1080/10986065.2021.1882363

Torres, L., Isa, I., Fajardo, P., Romero, S., Cárdenas, V., & Calderón, J. (2026). Do Home Digital Resources Bridge or Widen the Achievement Gap? Evidence from PISA 2022 in Latin America and the Caribbean. *Education Sciences*, *16*(10), 1709. https://doi.org/10.3390/educsci16101709

Tripo, J., & Debrenti, E. (2026). Mathematical development through non-digital instruction: a meta-analysis of primary and middle school education. *Frontiers in Education*, *11*. https://doi.org/10.3389/feduc.2026.1729068

Van Vo, D., & Csapó, B. (2023). Exploring Inductive Reasoning, Scientific Reasoning and Science Motivation, and Their Role in Predicting STEM Achievement Across Grade Levels. *International Journal of Science and Mathematics Education*, *21*(8), 2375-2398. https://doi.org/10.1007/s10763-022-10349-4

Vessonen, T., Hakkarainen, A., Väisänen, E., Laine, A., Aunio, P., & Gagnon, J. (2021). Differential Effects of Virtual and Concrete Manipulatives in a Fraction Intervention on Fourth and Fifth Grade Students' Fraction Skills. *Investigations in Mathematics Learning*, *13*(4), 323-337. https://doi.org/10.1080/19477503.2021.1982586

Webb, M. (2023). Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics/Revue canadienne d'économique*, *56*(3), 839-858. https://doi.org/10.1111/caje.12661

Wilson, S., Freeman, B., & Hedberg, E. (2026). Empirical Benchmarks for Effect Size Interpretation and Study Planning with Social and Behavioral Outcomes. *Journal of Research on Educational Effectiveness*, *19*(1), 129-154. https://doi.org/10.1080/19345747.2024.2427767

Wong, T., & Morsanyi, K. (2023). The link between transitive reasoning and mathematics achievement in preadolescence: the role of relational processing and deductive reasoning. *Thinking & Reasoning*, *29*(4), 531-558. https://doi.org/10.1080/13546783.2022.2095031

Yihan, L., Cuong, T., Norbert, S., & Józsa, K. (2026). Digital Game-Based Learning in Primary Education: A Cluster-Randomized Controlled Trial on Knowledge, Skills, and Motivation. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-026-10028-1

Zapata, M., Marcelino, P., El-Hamamsy, L., & Martín, E. (2024). A Bebras Computational Thinking (ABC-Thinking) program for primary school: Evaluation using the competent computational thinking test. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12441-w

CAPÍTULO X

APRENDIZAJE AUTORREGULADO, COMPROMISO ACADÉMICO Y ANSIEDAD MATEMÁTICA EN ESTUDIANTES UNIVERSITARIOS DE PRIMER AÑO DEL PERÚ: UN MODELO DE ECUACIONES ESTRUCTURALES CON MEDIACIÓN

Romer Juvenal Javier Quijano

Universidad Nacional Hermilio Valdizán

Resumen

La ansiedad matemática condiciona la permanencia y el rendimiento de los estudiantes universitarios de primer año, pero se desconoce si el aprendizaje autorregulado la reduce solo de forma directa o también mediante el compromiso académico con las matemáticas. Este estudio examinó, con un modelo de ecuaciones estructurales basado en covarianzas, las relaciones entre ambos procesos y la ansiedad matemática en estudiantes de primer año de universidades públicas y privadas del Perú durante 2025. Con un diseño cuantitativo, transversal y explicativo, se evaluó a 375 estudiantes seleccionados por conveniencia mediante escalas breves de tres ítems por constructo. El modelo de medición mostró cargas, fiabilidad y varianza extraída adecuadas, validez discriminante según la razón HTMT e invarianza entre mujeres y hombres. El modelo hipotetizado presentó un ajuste excelente y superó al modelo alternativo de un factor. El aprendizaje autorregulado se asoció positivamente con el compromiso ($\beta = .790$), el compromiso se asoció negativamente con la ansiedad ($\beta = -.260$) y la autorregulación mantuvo una asociación directa negativa con la ansiedad ($\beta = -.393$). El efecto indirecto (−.205; IC 95 % [−.377, −.019]) representó el 34.3 % del efecto total, lo que indica una mediación parcial, estable en los análisis de robustez. Los resultados sugieren que las intervenciones de acompañamiento en matemáticas serían más prometedoras si combinan el entrenamiento en autorregulación con el fomento del compromiso con la disciplina.

Palabras clave: aprendizaje autorregulado; compromiso académico; ansiedad matemática; estudiantes universitarios; ecuaciones estructurales; mediación

Abstract

Math anxiety shapes the persistence and performance of first-year university students, yet it remains unclear whether self-regulated learning reduces it only directly or also through academic engagement with mathematics. This study

used covariance-based structural equation modeling to examine the relations among both processes and math anxiety in first-year students from public and private universities in Peru during 2025. With a quantitative, cross-sectional, explanatory design, 375 students selected by convenience completed brief three-item scales for each construct. The measurement model showed adequate loadings, reliability, and average variance extracted, discriminant validity according to the HTMT ratio, and invariance between women and men. The hypothesized model showed excellent fit and outperformed the alternative one-factor model. Self-regulated learning was positively associated with engagement (= .790), engagement was negatively associated with math anxiety (β = −.260), and self-regulation retained a negative direct association with anxiety (β = −.393). The indirect effect (−.205; 95% CI [−.377, −.019]) accounted for 34.3% of the total effect, indicating partial mediation that remained stable across robustness analyses. The results suggest that academic support interventions in mathematics may be more promising when they combine self-regulation training with the promotion of engagement with the discipline.

Keywords: *self-regulated learning; academic engagement; math anxiety; university students; structural equation modeling; mediation*

10.1 INTRODUCCIÓN

La transición a la universidad es uno de los momentos en que la relación de los estudiantes con las matemáticas se vuelve más frágil: los cursos de primer año exigen un trabajo autónomo sostenido, evaluaciones frecuentes de alto peso y un ritmo de contenidos que ya no está mediado por la estructura escolar. En ese escenario, la ansiedad matemática —la respuesta de tensión, preocupación y evitación ante situaciones que implican manipular números o resolver problemas matemáticos— deja de ser un rasgo marginal y pasa a condicionar la permanencia y el rendimiento en carreras con base cuantitativa. La evidencia metaanalítica sitúa su asociación con el desempeño matemático en torno a r = −.28 (Barroso et al., 2021), y la síntesis de los estudios longitudinales examina la estabilidad de la ansiedad matemática y de esa relación a lo largo del

tiempo (Pelegrina et al., 2026). Los modelos estructurales con estudiantes universitarios vinculan la ansiedad matemática con la autorregulación y la autoeficacia (Zientek y Lane, 2026). En los estudiantes de primer año, la ansiedad matemática predice la evitación de los cursos de ciencias, tecnología, ingeniería y matemáticas y un menor rendimiento a lo largo de la carrera, con independencia de la habilidad matemática (Daker et al., 2021), y su nivel difiere entre las áreas de conocimiento universitarias (Zamora et al., 2023).

Frente a esa vulnerabilidad afectiva, el aprendizaje autorregulado —la capacidad de planificar, supervisar y sostener el propio esfuerzo de estudio (Zimmerman, 2002)— es uno de los recursos más investigados y más susceptibles de intervención: los programas de entrenamiento en autorregulación mejoran el rendimiento, las estrategias y la motivación de los estudiantes universitarios, con un tamaño del efecto global de g = .36 (Theobald, 2021). La autorregulación es además una de las variables que explican el rendimiento de los universitarios de primer año y una competencia que conviene promover desde el ingreso (Gálvez et al., 2024), y la estructura de sus competencias en estudiantes universitarios se ha examinado mediante análisis factorial (Mucha et al., 2025). Sin embargo, la literatura ha estudiado la autorregulación, el compromiso y la ansiedad casi siempre por pares. Una revisión sistemática de 28 estudios asocia los niveles altos de autorregulación con una menor ansiedad matemática (Harahap et al., 2025), resultado que los modelos estructurales con universitarios respaldan (Zientek y Lane, 2026), y la autorregulación se vincula con el compromiso en estudiantes de secundaria y universitarios (Dai et al., 2022; Shi y Li, 2026). Lo que sigue sin establecerse es el mecanismo que une los tres procesos: si la autorregulación reduce la ansiedad solo de manera directa o también porque genera una implicación sostenida con la disciplina, que a su vez modifica la valoración emocional de las tareas matemáticas.

Esta pregunta es relevante más allá de cualquier sistema universitario concreto. Desde la teoría del control-valor de las emociones de logro (Pekrun, 2024), la ansiedad surge cuando el estudiante percibe bajo control

sobre una tarea que valora; la autorregulación y el compromiso son, precisamente, las dos vías por las que el estudiante acumula experiencias de control. Contrastar si el compromiso transmite parte del efecto de la autorregulación sobre la ansiedad aporta una prueba directa de esa cadena teórica en la etapa de ingreso a la universidad, que es donde las intervenciones de acompañamiento académico se concentran en todo el mundo. El estudio se realiza con estudiantes universitarios de primer año del Perú, un contexto de expansión de la matrícula y de alta heterogeneidad en la preparación matemática previa, que funciona aquí como escenario de prueba de un modelo de alcance general. En el Perú, la investigación reciente ha descrito el compromiso académico de los universitarios (Estrada y Paricahua, 2023) y ha examinado la autorregulación en los entornos virtuales de la universidad pública (Gamarra et al., 2026), pero no ha contrastado el mecanismo que une los tres procesos.

El objetivo fue examinar, mediante un modelo de ecuaciones estructurales basado en covarianzas, las relaciones entre el aprendizaje autorregulado, el compromiso académico con las matemáticas y la ansiedad matemática en estudiantes universitarios de primer año del Perú durante el año académico 2025. La pregunta de investigación fue: ¿en qué medida el aprendizaje autorregulado se asocia con una menor ansiedad matemática de manera directa y a través del compromiso académico con las matemáticas? Para responderla se contrastan cuatro hipótesis, fundamentadas en el apartado siguiente:

1) H1, el aprendizaje autorregulado se relaciona positivamente con el compromiso académico con las matemáticas;

2) H2, el compromiso académico se relaciona negativamente con la ansiedad matemática;

3) H3, el aprendizaje autorregulado se relaciona negativamente con la ansiedad matemática; y

4) H4, el aprendizaje autorregulado se relaciona con una menor ansiedad matemática de manera indirecta, a través del compromiso académico con las matemáticas.

10.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

El modelo cíclico del aprendizaje autorregulado describe tres fases —anticipación, ejecución y autorreflexión— en las que el estudiante fija metas, supervisa su comprensión y ajusta sus estrategias (Zimmerman, 2002). Un estudiante que planifica y supervisa su estudio invierte más tiempo y atención en la disciplina y persiste ante la dificultad, que son las manifestaciones conductuales y cognitivas del compromiso académico (Veiga et al., 2026). La evidencia con estudiantes de distintos niveles respalda esta relación positiva (Dai et al., 2022; Shi y Li, 2026). De ahí la primera hipótesis.

1) **H1.** El aprendizaje autorregulado se relaciona positivamente con el compromiso académico con las matemáticas.

   La teoría del control-valor (Pekrun, 2024) sostiene que las emociones de logro dependen de las valoraciones de control y de valor que el estudiante hace de la tarea. El compromiso sostenido produce experiencias repetidas de dominio que elevan el control percibido, y los estudios longitudinales muestran relaciones recíprocas entre el control percibido y la ansiedad en el aula de matemáticas (Putwain y Wood, 2023) y entre las valoraciones cognitivas, las emociones de logro y el rendimiento en matemáticas (Forsblom et al., 2022). En estudiantes de distintos sistemas educativos, el compromiso conductual se asocia con menor ansiedad matemática (Bergqvist, 2026; Shimizu, 2025).

2) **H2.** El compromiso académico con las matemáticas se relaciona negativamente con la ansiedad matemática.

   Con independencia del compromiso, la autorregulación proporciona al estudiante herramientas para anticipar las demandas de una evaluación, detectar a tiempo las lagunas de comprensión y regular sus

respuestas afectivas, lo que reduce la percepción de amenaza (Klee et al., 2022; Szucs y Toffalini, 2023). Los estudios recientes documentan asociaciones negativas entre autorregulación y ansiedad matemática (Harahap et al., 2025; Zientek y Lane, 2026).

3) **H3.** El aprendizaje autorregulado se relaciona negativamente con la ansiedad matemática.

   Si la autorregulación eleva el compromiso (H1) y el compromiso reduce la ansiedad (H2), parte del efecto de la autorregulación sobre la ansiedad debe transmitirse a través del compromiso. Este planteamiento traduce la secuencia del control-valor —recursos de regulación, experiencias de dominio y valoración emocional— en una hipótesis de mediación contrastable.

4) **H4**. El aprendizaje autorregulado se relaciona negativamente con la ansiedad matemática de manera indirecta, a través del compromiso académico con las matemáticas.

   El sexo y el tipo de gestión de la universidad (pública o privada) se incorporan como covariables de control en los análisis de robustez, sin rol estructural en el modelo.

10.3 MATERIALES Y MÉTODOS

10.3.1 Diseño y participantes

El estudio siguió un diseño cuantitativo, no experimental, transversal y explicativo, con el estudiante como unidad de análisis. La población estuvo constituida por estudiantes matriculados en el primer año de universidades públicas y privadas del Perú durante el año académico 2025. El tamaño muestral se fijó a priori con la prueba de ajuste cercano del RMSEA (MacCallum et al., 1996) para el modelo de tres factores con tres indicadores cada uno ($p = 9$, $q = 21$ parámetros libres, $gl = 24$; H0: $\varepsilon = .05$ frente a H1: $\varepsilon = .08$; $\alpha = .05$): $N = 375$ alcanzó una potencia de

.801 (.7998 con N = 374), el menor tamaño que cumple el criterio de .80. Un segundo criterio, la prueba F de un coeficiente en la ecuación de la ansiedad matemática (dos predictores), se anclo al tamaño de efecto de la asociación entre ansiedad y rendimiento matemático documentada en la literatura metaanalítica ($r = -.28$; $f^2 = .085$) y exigió N = 95 para una potencia de .803; con N = 375 la potencia supera .999. Rige, por tanto, el criterio del RMSEA. El muestreo fue no probabilístico por conveniencia, de modo que no se afirma representatividad nacional. La distribución por sexo y por tipo de gestión de la universidad fue equilibrada (50.4 % y 49.6 % en las dos categorías de cada variable). Los criterios de inclusión fueron estar matriculado en el primer año, otorgar consentimiento informado y completar los tres instrumentos; los cuestionarios incompletos se excluyeron.

10.3.2 Instrumentos

Los tres constructos se midieron con escalas reflectivas de tres ítems y respuesta tipo Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo), cuya operacionalización resume la Tabla 10-1. La escala de aprendizaje autorregulado en matemáticas adaptó las subescalas de autorregulación metacognitiva y de regulación del esfuerzo del cuestionario validado con universitarios peruanos (Rocha et al., 2023); la de compromiso académico con las matemáticas adaptó la escala breve de compromiso en actividades de aprendizaje (Veiga et al., 2026), y la medición del compromiso en universitarios de Ecuador y Perú dispone de evidencia psicométrica (Arevalo et al., 2026); y la de ansiedad matemática adaptó la Abbreviated Math Anxiety Scale en su validación peruana (Quiroz et al., 2026), en línea con las escalas breves para universitarios (Núñez y Guilera, 2023). En cada caso se seleccionaron tres ítems representativos de las dimensiones del instrumento de origen y se adaptó su redacción al curso universitario de matemáticas. Las propiedades psicométricas en esta muestra se reportan en Resultados.

Tabla 10-1

Operacionalización de las variables del estudio

Constructo	Rol en el modelo	Definición operativa	Instrumento	Ítems	Escala	
Aprendizaje autorregulado	Exógena latente	Planificación, monitoreo y regulación del esfuerzo en el estudio de las matemáticas	Escala de aprendizaje autorregulado en matemáticas, adaptada de Rocha et al. (2023)	3	Likert de 5 puntos	
Compromiso académico con las matemáticas	Mediadora latente	Implicación conductual, cognitiva y emocional en las actividades de los cursos de matemáticas	Escala de compromiso académico con las matemáticas, adaptada de Veiga et al. (2026)			
Ansiedad matemática	Endógena latente	Tensión y preocupación ante la evaluación y el aprendizaje de las matemáticas	Escala abreviada de ansiedad matemática, adaptada de Quiroz et al. (2026)			
Sexo	Covariable de control	Sexo declarado por el participante	Registro sociodemográfico	1	Dicotómica	(mujer, hombre)
Tipo de gestión de la universidad		Régimen de gestión de la universidad de matrícula	Registro institucional			(pública, privada)

Nota: Las covariables se incorporan en los análisis de robustez (Tabla 10-9); el modelo estructural incluye los tres constructos latentes.

10.3.3 Procedimiento

Los cuestionarios se aplicaron durante el segundo semestre del año académico 2025, en horario de clase. Antes de responder, los estudiantes recibieron información sobre los objetivos del estudio, el carácter voluntario y anónimo de su participación y la confidencialidad de sus respuestas.

10.3.4 Análisis de datos

El análisis se desarrolló en dos etapas. La primera evaluó el modelo de medición mediante análisis factorial confirmatorio: cargas estandarizadas (λ), fiabilidad (α de Cronbach y ω), varianza extraída media (AVE) y validez discriminante por la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015) y por el criterio de Fornell y Larcker (1981). La invarianza entre mujeres y hombres se examinó con modelos configural, métrico y escalar anidados, con el criterio ΔCFI ≥ −.010 (Cheung y Rensvold, 2002) y el ΔRMSEA como complemento. La segunda etapa estimó por máxima verosimilitud el modelo estructural basado en covarianzas, con las rutas a (aprendizaje autorregulado → compromiso), b (compromiso → ansiedad) y c′ (aprendizaje autorregulado → ansiedad); el efecto indirecto a × b y el total se estimaron con intervalos de confianza al 95 % por bootstrap de 5000 réplicas, y se consideró distinto de cero el efecto cuyo intervalo excluía el cero. El ajuste global se evaluó con χ^2, χ^2/gl, CFI, TLI, RMSEA y SRMR, frente a un modelo alternativo de un solo factor, con valores de referencia de CFI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu y Bentler, 1999). Como análisis de robustez se estimó una regresión jerárquica por mínimos cuadrados ordinarios sobre las puntuaciones compuestas, con errores estándar HC3, que introdujo primero las covariables, después el aprendizaje autorregulado y por último el compromiso; la estabilidad del efecto directo de la autorregulación se evaluó con jackknife, submuestreos del 80 % y del 60 % y 1000 remuestreos bootstrap. Los análisis se realizaron en Python con la librería semopy.

10.4 RESULTADOS

10.4.1 Modelo de medición

Las nueve cargas factoriales estandarizadas oscilaron entre .639 y .847, todas por encima de .50; la fiabilidad (α y ω entre .760 y .840) y la AVE (entre .518 y .637) superaron los umbrales establecidos en los tres constructos (Tabla 10-2).

Tabla 10-2

Modelo de medición: propiedades psicométricas por ítem

Constructo	Ítem	λ	R^2	α	ω	AVE
Aprendizaje autorregulado	Antes de estudiar matemáticas planifico las metas y los pasos de cada sesión	.847	.717	.760	.760	.518
	Mientras resuelvo problemas compruebo si estoy comprendiendo y ajusto mi estrategia	.639	.408	.760	.760	.518
	Cuando una tarea de matemáticas se vuelve difícil mantengo el esfuerzo hasta terminarla	.655	.429	.760	.760	.518
Compromiso académico con las matemáticas	Participo activamente en las clases y actividades de matemáticas	.749	.561	.840	.840	.637
	Dedico tiempo a profundizar los contenidos de matemáticas más allá de lo exigido	.818	.669	.840	.840	.637
	Me siento implicado e interesado en el trabajo de los cursos de matemáticas	.825	.681	.840	.840	.637
Ansiedad matemática	Me pongo nervioso cuando pienso en un examen de matemáticas	.676	.457	.786	.787	.553
	Siento tensión cuando debo resolver problemas de matemáticas frente a otros	.798	.637	.786	.787	.553
	Me preocupa no entender los nuevos temas de matemáticas que se explican en clase	.752	.566	.786	.787	.553

Nota: λ = carga factorial estandarizada; R^2 = correlación múltiple al cuadrado del ítem; α = alfa de Cronbach; ω = omega; AVE = varianza extraída media. Criterios: λ ≥ .50; α y ω ≥ .70; AVE ≥ .50.

La razón HTMT se mantuvo por debajo del umbral estricto de .85 en los tres pares de constructos, con el valor más alto entre el aprendizaje autorregulado y el compromiso académico. El criterio de Fornell y Larcker se cumplió para la ansiedad matemática y para el compromiso, pero no para el aprendizaje autorregulado, cuya raíz cuadrada de la AVE quedó por debajo de su correlación con el compromiso (Tabla 10-3). Esta proximidad entre ambos constructos se retoma en la Discusión.

Tabla 10-3

Validez discriminante según la razón HTMT y el criterio de Fornell y Larcker

Criterio	Par de constructos o constructo	Valor	Referencia	Cumple
HTMT	Aprendizaje autorregulado — Ansiedad matemática	.515	< .85	Sí
	Aprendizaje autorregulado — Compromiso académico con las matemáticas	.794	< .85	Sí
	Ansiedad matemática — Compromiso académico con las matemáticas	.527	< .85	Sí
Fornell y Larcker	Aprendizaje autorregulado	√AVE = .720; r máx = .784	√AVE > r máx	No
	Ansiedad matemática	√AVE = .744; r máx = .592	√AVE > r máx	Sí
	Compromiso académico con las matemáticas	√AVE = .798; r máx = .784	√AVE > r máx	Sí

Nota: HTMT = razón heterorrasgo-monorrasgo (Henseler et al., 2015); √AVE = raíz cuadrada de la varianza extraída media; r máx = mayor correlación del constructo con los demás constructos del modelo (Fornell y Larcker, 1981).

La estructura de medición fue equivalente entre mujeres y hombres según el criterio del ΔCFI, tanto en el paso del modelo configural al métrico como en el paso al escalar; el RMSEA aumentó en el paso métrico, partiendo de un modelo configural con RMSEA = .000, y no varió en el paso escalar (Tabla 10-4).

Tabla 10-4

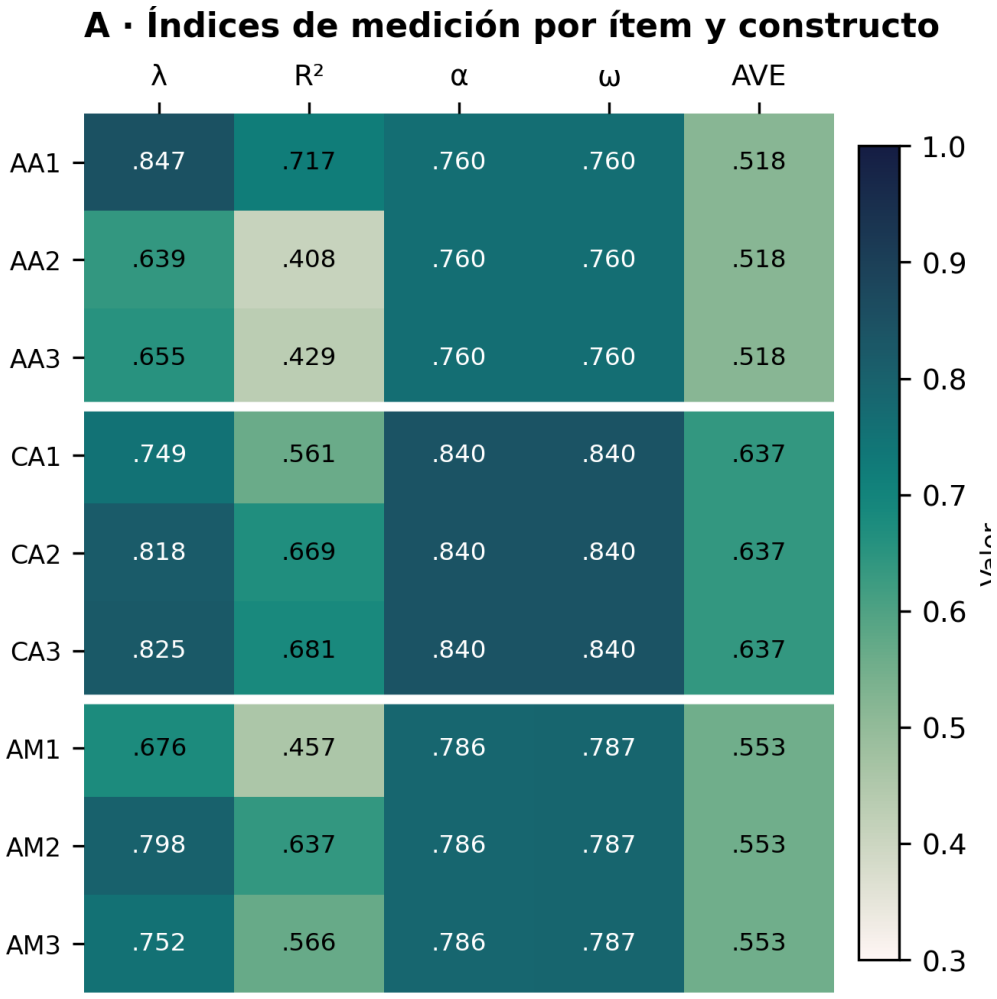
Invarianza de medida entre mujeres y varones

Modelo	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Invarianza (ΔCFI ≥ −.010)
Configural	1.000	1.000	.000	—	—	—
Métrico	.998	.997	.018	−.002	.018	Sí
Escalar	.998	.997	.018	.000	.000	Sí

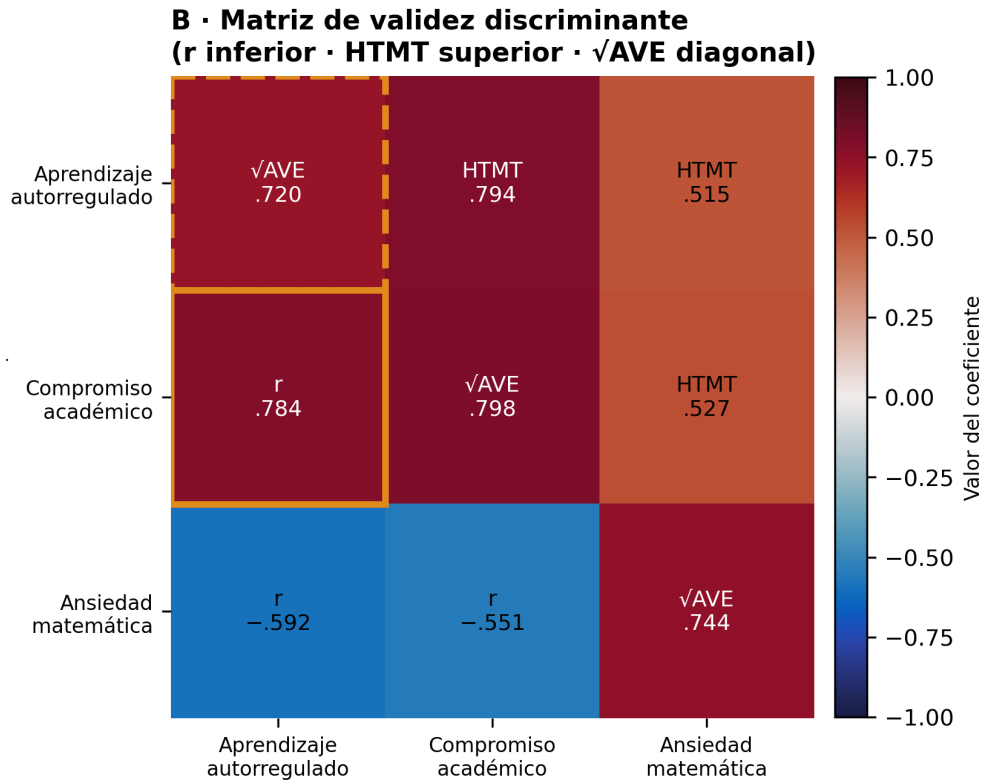
Nota: Modelos anidados estimados en los grupos de mujeres y hombres. Criterio de decisión: ΔCFI ≥ −.010 (Cheung y Rensvold, 2002); ΔRMSEA se reporta de forma complementaria..

Figura 10-1

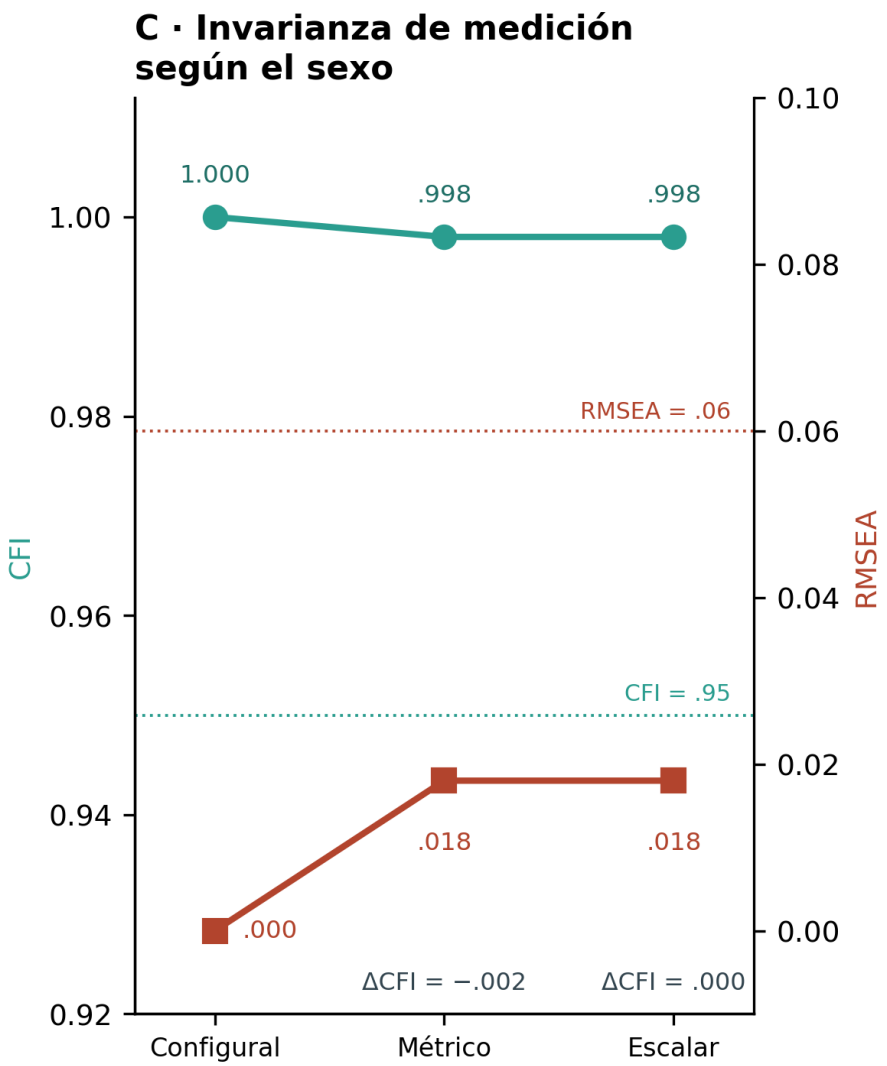
Evidencia de validez del modelo de medición



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel A: carga factorial estandarizada (λ) y comunalidad (R^2) de cada ítem, con el α de Cronbach, el ω y la AVE de su constructo (AA, aprendizaje autorregulado; CA, compromiso académico con las matemáticas; AM, ansiedad matemática); λ = .639 a .847, R^2 = .408 a .717, α = .760 a .840, AVE = .518 a .637.



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel B: matriz de validez discriminante con las correlaciones entre constructos bajo la diagonal (–.592, –.551 y .784), la razón HTMT sobre la diagonal (.515, .527 y .794; todas inferiores a .85) y la raíz cuadrada de la AVE en la diagonal (.720, .798 y .744); el recuadro continuo marca la correlación entre el aprendizaje autorregulado y el compromiso académico (.784), que supera la √AVE del aprendizaje autorregulado (.720, recuadro discontinuo), único caso en que no se cumple el criterio de Fornell y Larcker.



Nota. N = 375; análisis factorial confirmatorio por máxima verosimilitud. Panel C: CFI (eje izquierdo) y RMSEA (eje derecho) de los modelos configural, métrico y escalar estimados en mujeres y hombres, con ΔCFI = −.002 y .000 en los pasos métrico y escalar; líneas punteadas en CFI = .95 y RMSEA = .06 (Hu y Bentler, 1999).

Los modelos restringidos mantuvieron índices de ajuste excelentes. La Figura 10-1 sintetiza la evidencia de validez del modelo de medición.

10.4.2 Estadísticos descriptivos y correlaciones

Las puntuaciones de los tres constructos presentaron asimetría y curtosis cercanas a cero (Tabla 10-5). El aprendizaje autorregulado se correlacionó positivamente con el compromiso académico y negativamente con la ansiedad matemática, y el compromiso se correlacionó negativamente con la ansiedad, todas con $p < .001$. Ninguna de las dos covariables se asoció significativamente con los constructos.

Tabla 10-5

Estadísticos descriptivos y correlaciones

	Variable	M	DE	Asimetría	Curtosis	1	2	3	4
1.	Aprendizaje autorregulado	3.00	0.666	0.089	0.078	—			
2.	Ansiedad matemática	3.00	0.665	−0.155	0.127	−.592*** [−.654, −.522]	—		
3.	Compromiso académico con las matemáticas	3.00	0.667	0.045	−0.231	.784*** [.741, .820]	−.551*** [−.618, −.476]	—	
4.	Sexo	0.504	0.501	−0.016	−2.000	.063 [−.039, .163]	.072 [−.029, .172]	.061 [−.040, .162]	—
5.	Tipo de gestión de la universidad	0.504	0.501	−0.016	−2.000	−.052 [−.152, .050]	.030 [−.072, .130]	.009 [−.093, .110]	.019 [−.083, .120]

Nota: N = 375. M y DE de los constructos en la escala Likert de cinco puntos; las covariables están codificadas 0-1. Entre corchetes, intervalos de confianza al 95 % por la transformación Z de Fisher. ***p < .001.

10.4.3 Modelo estructural

El modelo hipotetizado mostró un ajuste excelente y superó con claridad al modelo alternativo de un factor, cuyo ajuste fue insuficiente en todos los índices (Tabla 10-6). El modelo de medición de la primera etapa obtuvo un SRMR = .034.

Tabla 10-6

Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor

Índice	Modelo		Valor de referencia
	hipotetizado (tres factores)	alternativo (un factor)	
χ^2	19.83	287.56	—
gl	24	27	—
p (χ^2)	.707	< .001	> .05
χ^2/gl	0.83	10.65	—
CFI	1.000	.808	≥ .95
TLI	1.000	.744	—
RMSEA	.000	.161	≤ .06
SRMR	.021	.307	≤ .08

Nota: Estimación por máxima verosimilitud. Valores de referencia de CFI, RMSEA y SRMR según Hu y Bentler (1999).

Las tres rutas directas fueron significativas y en la dirección esperada: el aprendizaje autorregulado se asoció positivamente con el compromiso académico (ruta a), el compromiso se asoció negativamente con la ansiedad matemática (ruta b) y la autorregulación mantuvo una asociación negativa directa con la ansiedad (ruta c′), como muestra la Tabla 10-7.

Tabla 10-7

Rutas estructurales estandarizadas del modelo CB-SEM y contraste de hipótesis

Ruta	Hipótesis	β	EE	p
Aprendizaje autorregulado → Compromiso académico con las matemáticas	H1	.790***	.063	< .001
Compromiso académico con las matemáticas → Ansiedad matemática	H2	−.260*	.104	.024
Aprendizaje autorregulado → Ansiedad matemática	H3	−.393**	.099	.001
Aprendizaje autorregulado → Compromiso académico → Ansiedad matemática (indirecto)	H4	−.205*	—	.034

Nota: β = coeficiente estandarizado; EE = error estándar. El efecto indirecto se evalúa con su intervalo de confianza bootstrap en la Tabla 8. *p < .05. **p < .01. ***p < .001.

La descomposición de efectos confirmó la mediación: el intervalo del efecto indirecto excluyó el cero y el efecto directo se mantuvo distinto de cero, lo que corresponde a una mediación parcial, con un 34.3 % del efecto total atribuible al componente indirecto y un 65.7 % al directo (Tabla 10-8). En consecuencia, las hipótesis H1, H2, H3 y H4 fueron soportadas. La Figura 10-2 integra el modelo estimado, la descomposición del efecto y los índices de ajuste.

Tabla 10-8

Descomposición de efectos del aprendizaje autorregulado sobre la ansiedad matemática

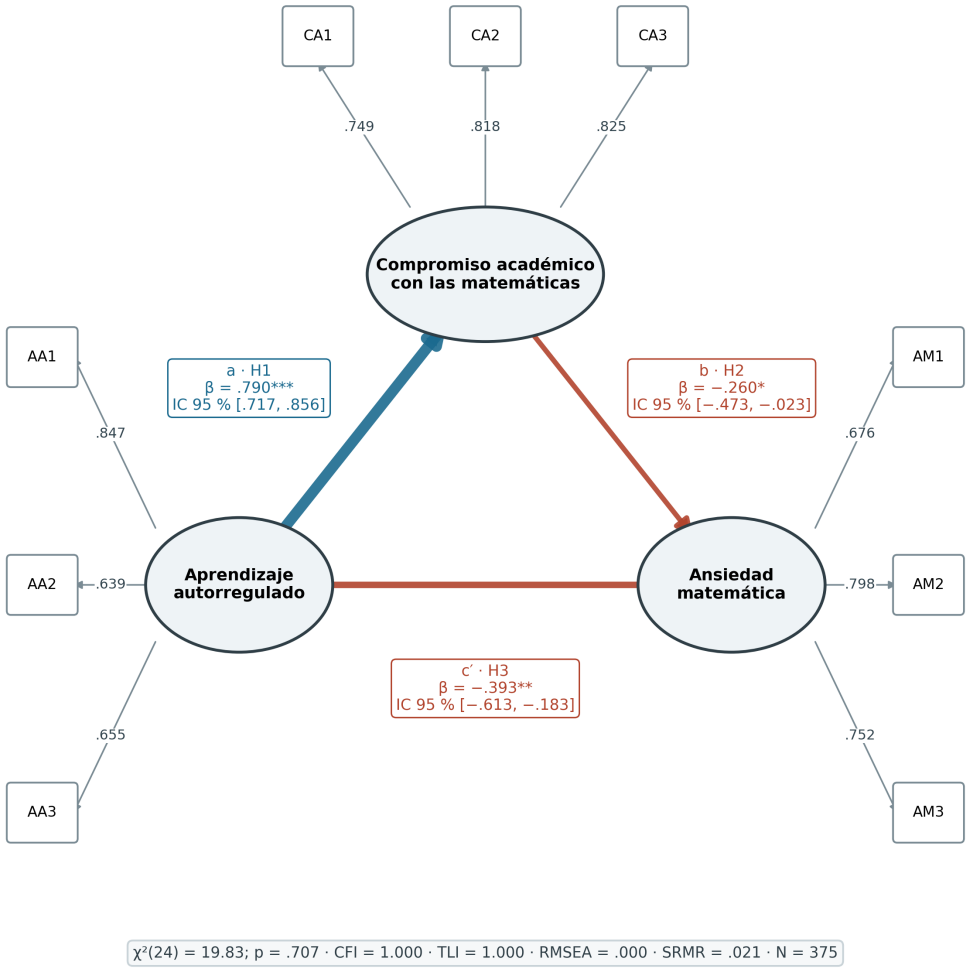
	Efecto	Coeficiente	IC 95 %		Distinto de cero
			inferior	superior	
a:	Aprendizaje autorregulado → Compromiso académico	.790	.717	.856	Sí
b:	Compromiso académico → Ansiedad matemática	−.260	−.473	−.023	Sí
c′:	efecto directo	−.393	−.613	−.183	Sí
a × b:	efecto indirecto	−.205	−.377	−.019	Sí
c:	efecto total (c′ + a × b)	−.599	−.682	−.507	Sí

Nota: Coeficientes estandarizados con intervalos de confianza al 95 % por bootstrap de 5000 réplicas. El efecto indirecto representa el 34.3 % del efecto total y el efecto directo, el 65.7 %.

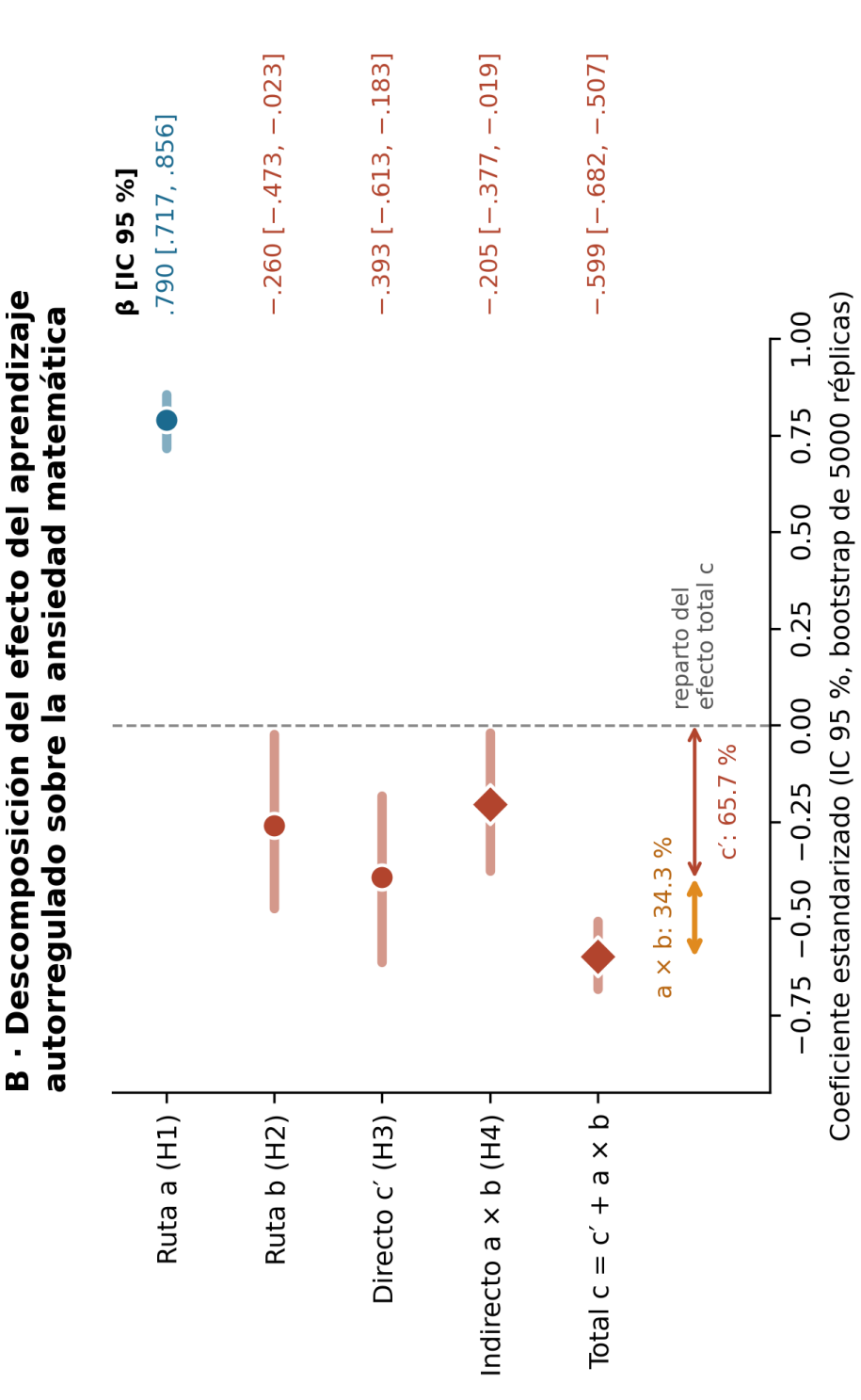
Figura 10-2

Modelo estructural estimado, descomposición del efecto e índices de ajuste

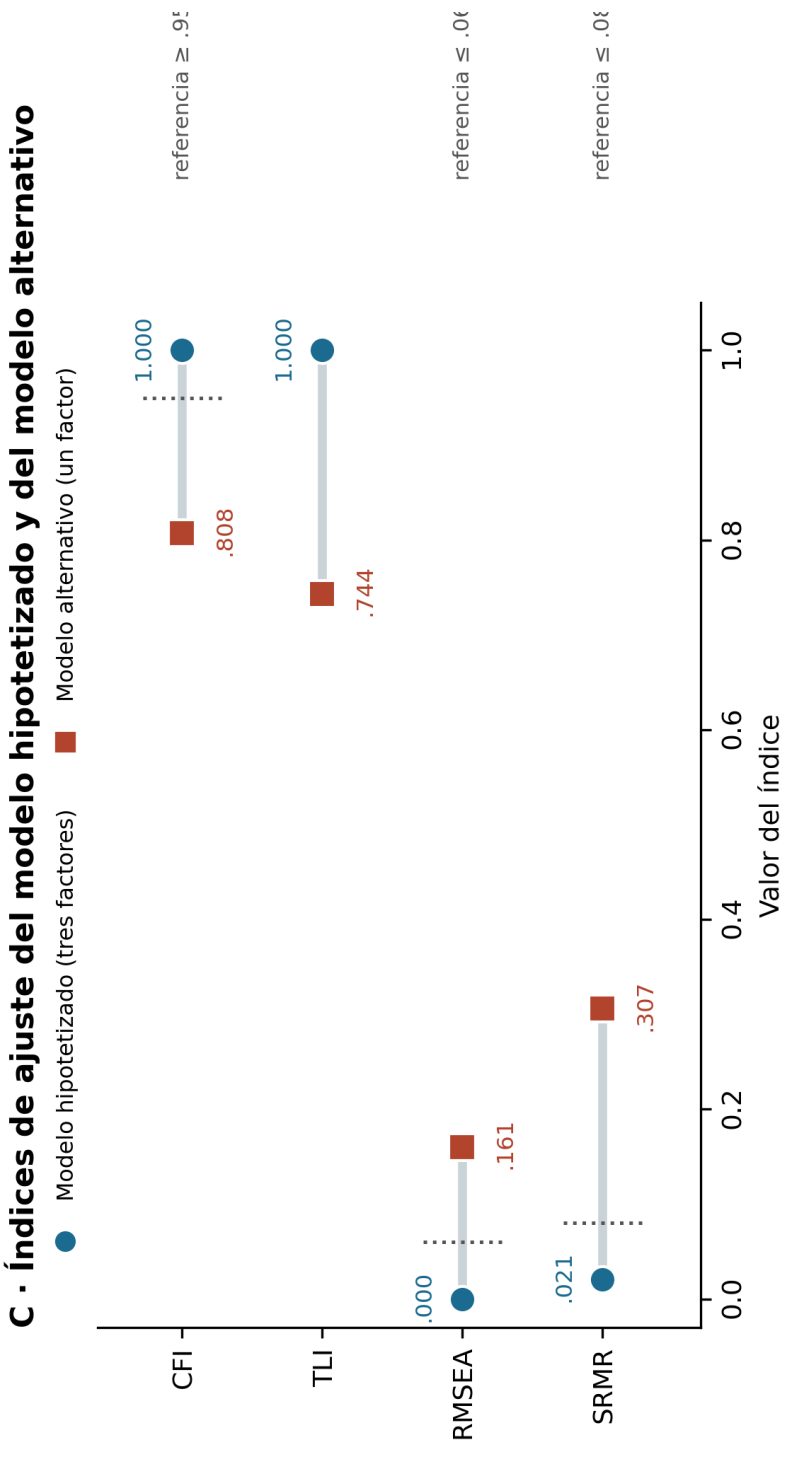
A · Modelo estructural estimado (CB-SEM, máxima verosimilitud)



Nota. N = 375; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel A: constructos latentes (elipses) con sus tres indicadores (AA1-AA3, aprendizaje autorregulado; CA1-CA3, compromiso académico con las matemáticas; AM1-AM3, ansiedad matemática) y sus cargas estandarizadas (.639 a .847); el grosor de cada ruta es proporcional a |β| y el color indica el signo (azul, positivo; rojo, negativo): a = .790, IC 95 % [.717, .856]; b = −.260, IC 95 % [−.473, −.023]; c′ = −.393, IC 95 % [−.613, −.183]; ajuste global $\chi^2(24) = 19.83$, p = .707, CFI = 1.000, TLI = 1.000, RMSEA = .000, SRMR = .021.



Nota. $N = 375$; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel B: coeficientes estandarizados de las rutas a, b y c′, del efecto indirecto a × b (−.205, IC 95 % [−.377, −.019]) y del efecto total c (−.599, IC 95 % [−.682, −.507]), con intervalos bootstrap de 5000 réplicas; las flechas inferiores reparten el efecto total entre el componente directo (65.7 %) y el indirecto (34.3 %).



Nota. N = 375; modelo de ecuaciones estructurales basado en covarianzas estimado por máxima verosimilitud. Panel C: CFI, TLI, RMSEA y SRMR del modelo hipotetizado de tres factores (círculos: 1.000, 1.000, .000 y .021) y del modelo alternativo de un factor (cuadrados: .808, .744, .161 y .307), con $\chi^2(24) = 19.83$ frente a $\chi^2(27) = 287.56$; líneas punteadas en los valores de referencia CFI = .95, RMSEA = .06 y SRMR = .08 (Hu y Bentler, 1999). *p < .05. **p < .01. ***p < .001.

10.4.4 Análisis de robustez

La regresión jerárquica sobre las puntuaciones compuestas reprodujo la dirección y la significación de las rutas del modelo estructural (Tabla 10-9). Las covariables solas no explicaron la ansiedad matemática; la entrada del aprendizaje autorregulado añadió la mayor parte de la varianza explicada y la del compromiso académico un incremento menor, y en el modelo final ambos predictores se asociaron negativamente con la ansiedad, con VIF ≤ 1.01. El sexo mostró una asociación positiva pequeña con la ansiedad una vez controlados los dos predictores, mientras que el tipo de gestión de la universidad no se asoció con ella. La Figura 10-3 resume estas relaciones.

Tabla 10-9

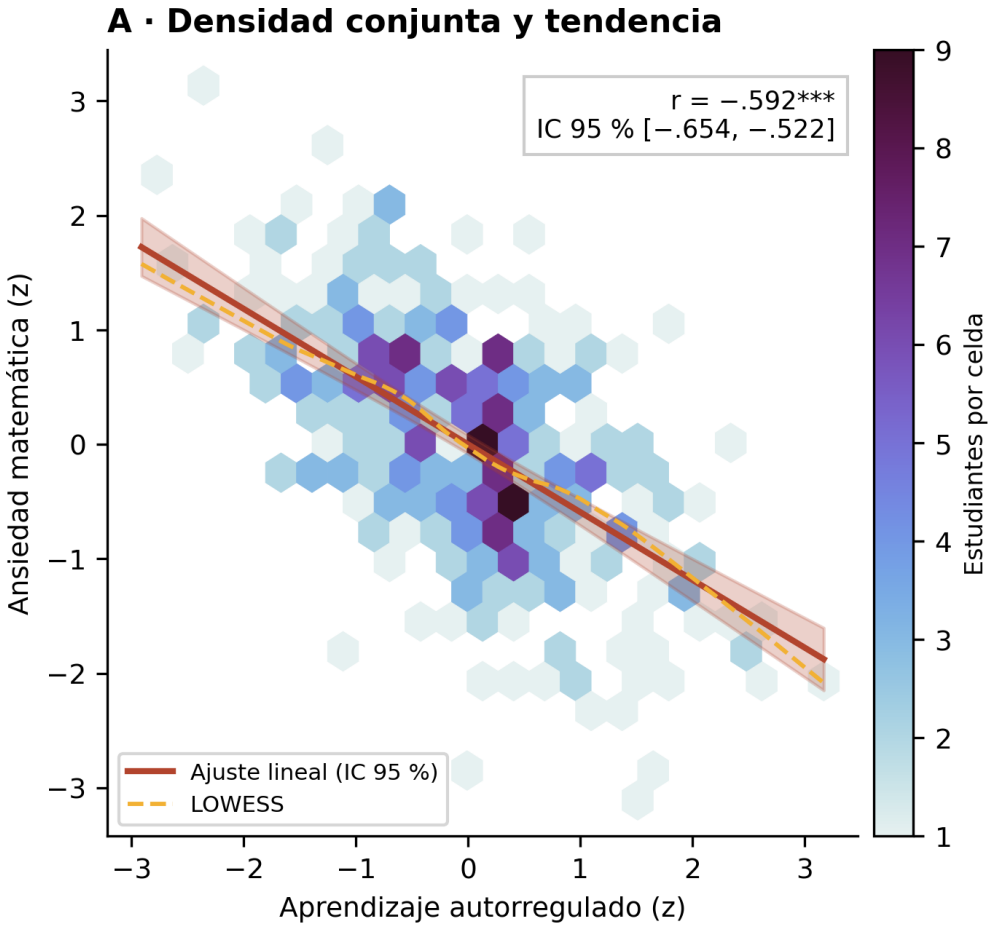
Regresión jerárquica de la ansiedad matemática sobre las covariables y los predictores (MCO con errores HC3)

Predictor o estadístico	M1	M2	M3
Sexo	0.072 (0.052)	0.110** (0.042)	0.113** (0.041)
Tipo de gestión de la universidad	0.028 (0.052)	–0.004 (0.042)	0.008 (0.041)
Aprendizaje autorregulado	—	–0.300*** (0.020)	–0.210*** (0.030)
Compromiso académico con las matemáticas	—	—	–0.212*** (0.056)
R^2	.006	.363	.383
R^2 ajustado	.001	.357	.376
ΔR^2	—	.357	.020
F (p)	1.13 (.324)	76.54 (< .001)	61.32 (< .001)

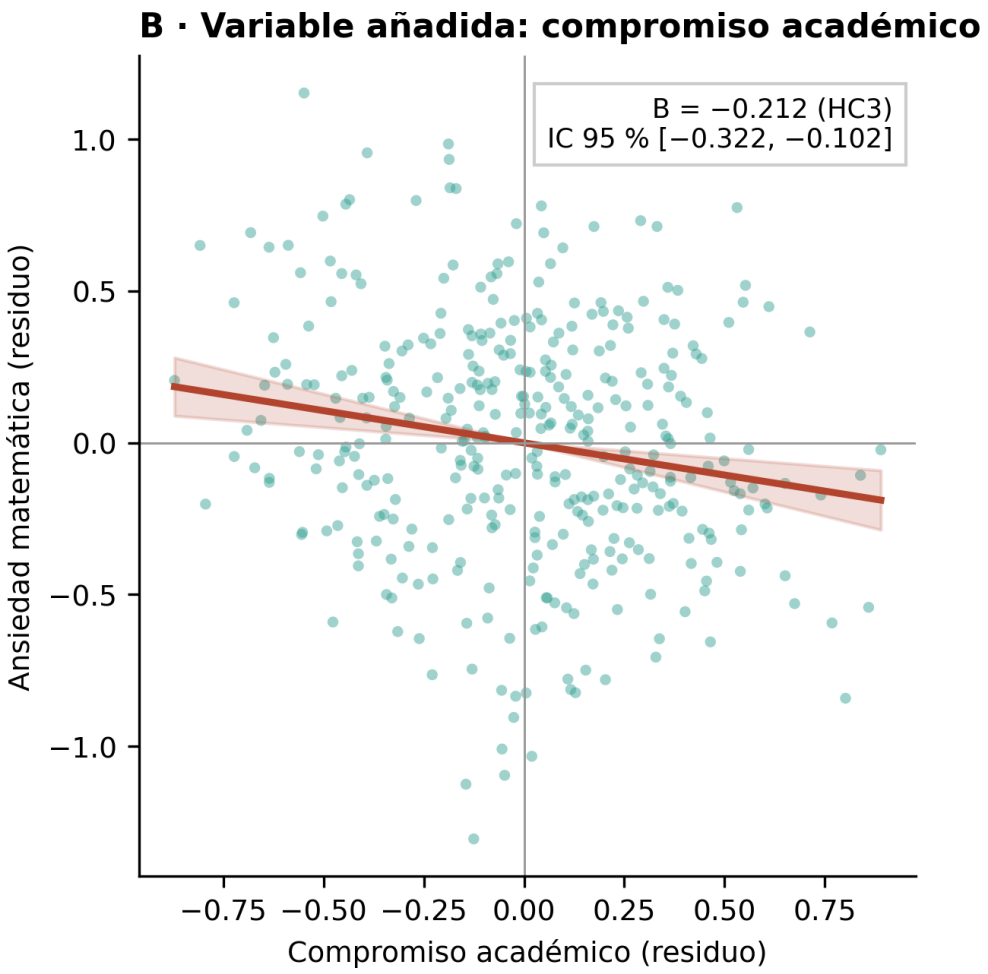
Nota: Coeficientes no estandarizados con errores estándar HC3 entre paréntesis, estimados sobre las puntuaciones compuestas de las escalas; N = 375. Al no declararse términos de interacción, un cuarto bloque reproduce exactamente el modelo M3 y se omite. **p < .01. ***p < .001.

Figura 10-2

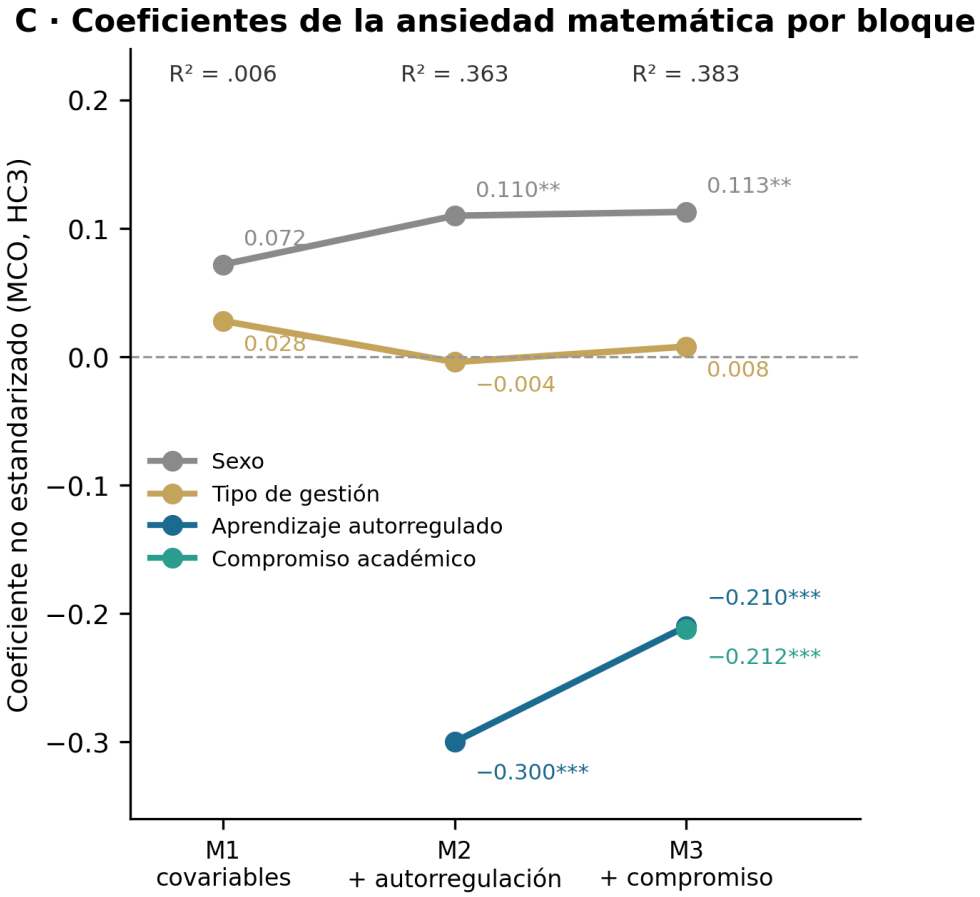
Relaciones observadas entre los constructos y coeficientes de la regresión jerárquica de la ansiedad matemática



Nota. N = 375; puntuaciones compuestas de las escalas. Panel A: densidad conjunta (hexágonos) del aprendizaje autorregulado y la ansiedad matemática en puntuaciones estandarizadas, con el ajuste lineal y su IC 95 % y una curva LOWESS; r = −.592, IC 95 % [−.654, −.522].



Nota. N = 375; puntuaciones compuestas de las escalas. Panel B: gráfico de variable añadida del compromiso académico con las matemáticas sobre la ansiedad matemática, una vez descontados el aprendizaje autorregulado, el sexo y el tipo de gestión de la universidad; la pendiente reproduce el coeficiente del modelo M3 de la Tabla 9, B = −0.212 (errores HC3), IC 95 % [−0.322, −0.102].



Nota. N = 375; puntuaciones compuestas de las escalas. Panel C: coeficientes no estandarizados de cada predictor en los tres bloques de la regresión jerárquica (Tabla 9): sexo 0.072, 0.110 y 0.113; tipo de gestión 0.028, –0.004 y 0.008; aprendizaje autorregulado –0.300 en M2 y –0.210 en M3; compromiso académico –0.212 en M3; R^2 = .006, .363 y .383. **p < .01. ***p < .001.

El efecto directo del aprendizaje autorregulado sobre la ansiedad matemática fue estable ante las perturbaciones de la muestra: conservó el signo y la significación en todas las réplicas de las cuatro pruebas, con coeficientes medios muy próximos al valor de la muestra completa (Tabla 10-10).

Tabla 10-10

Estabilidad del efecto del aprendizaje autorregulado sobre la ansiedad matemática ante perturbaciones de la muestra

Prueba	Réplicas	β medio	DE	IC 95 %	Signo estable (%)	Significativo (%)
Jackknife (exclusión de un caso)	375	−0.3005	0.0010	[−0.3026, −0.2984]	100 %	100 %
Submuestreo del 80 %	100	−0.3000	0.0087	[−0.3193, −0.2834]	100 %	100 %
Submuestreo del 60 %	100	−0.2991	0.0165	[−0.3362, −0.2725]	100 %	100 %
Remuestreo bootstrap	1000	−0.3011	0.0195	[−0.3397, −0.2629]	100 %	100 %

Nota: Coeficiente no estandarizado del aprendizaje autorregulado en la regresión de la ansiedad matemática sobre las puntuaciones compuestas con las covariables de control (modelo M2 de la Tabla 9; B = −0.3005 en la muestra completa, −0.300 con tres decimales). «Signo estable» y «significativo» indican el porcentaje de réplicas que conservan el signo negativo y $p < .05$.

10.5 DISCUSIÓN

El hallazgo más desconcertante no es la mediación en sí, sino que se sostenga entre dos constructos casi superpuestos: la autorregulación y el compromiso covarían de forma muy intensa ($r = .784$), la raíz cuadrada de la AVE de la autorregulación queda por debajo de esa correlación y solo la razón HTMT (.794) respalda su distinción. Aun así, el compromiso conservó una asociación propia con la ansiedad ($\beta = -.260$, $p = .024$) al controlar la autorregulación, y el efecto indirecto ($a \times b = -.205$, IC 95 %

[−.377, −.019]) representó un 34.3 % del efecto total. Las cuatro hipótesis recibieron apoyo, pero el límite superior del intervalo indirecto se acerca al cero, de modo que los datos son compatibles con una mediación parcial cuya magnitud se estima con imprecisión. El compromiso no parece, por tanto, un correlato accesorio de la autorregulación, aunque tampoco puede afirmarse que sea su mecanismo principal: el componente directo ($c' = -.393$) fue mayor que el indirecto.

Este patrón es coherente con la teoría del control-valor (Pekrun, 2024), que atribuye las emociones de logro a las valoraciones de control y de valor que el estudiante construye sobre la tarea. Que una parte del efecto de la autorregulación transite por el compromiso sugiere que disponer de estrategias de planificación y supervisión no basta: su relación con la ansiedad parecería depender también de que se traduzcan en implicación sostenida con la disciplina, donde se acumularían las experiencias de dominio. La mayor magnitud del componente directo es compatible con la idea de que la autorregulación aporta recursos para anticipar las demandas de la evaluación y modular la propia respuesta afectiva; los estudiantes con ansiedad matemática tienden a centrarse en no fracasar más que en dominar el contenido y a percibir un bajo control sobre los resultados (Klee et al., 2022). La evidencia longitudinal que enlaza valoraciones cognitivas, emociones de logro y rendimiento en matemáticas (Forsblom et al., 2022) ofrece un marco para leer la cadena propuesta, aunque aquí no se midieron el control ni el valor, de modo que la secuencia se infiere y no se observa.

Desde una perspectiva crítica, esa lectura tiene aristas que los datos no resuelven. Putwain y Wood (2023) hallaron relaciones bidireccionales entre mayor ansiedad y menor control y, si bien la ansiedad predijo menor valor, el valor no se relacionó con la ansiedad posterior; Szucs y Toffalini (2023) encontraron que la ansiedad se asociaba con menor control percibido y menor expectativa de éxito y, de forma sorprendente, que una mayor valoración subjetiva de las matemáticas se acompañaba de mayor ansiedad para niveles similares de control. Si el valor otorgado a la disciplina no reduce la ansiedad y esta puede erosionar el control,

la dirección asumida por el modelo, de la regulación al compromiso y de este a la ansiedad, no es la única compatible con una covariación transversal: la ansiedad podría, a su vez, deprimir el compromiso. En esa línea, Cahyawati et al. (2023) observaron que la ansiedad matemática afectaba de manera indirecta al aprendizaje autorregulado a través del autoconcepto, un orden inverso al postulado aquí, y que los estudiantes con baja ansiedad tendían a mostrar una autorregulación alta. La escala de compromiso, además, fusiona en tres ítems las implicaciones conductual, cognitiva y emocional, por lo que no permite saber cuál sostiene el efecto; Bergqvist (2026) indica que la ansiedad matemática se asocia sobre todo con el desenganche cognitivo, lo que sugiere que el componente cognitivo podría concentrar buena parte de la mediación observada, conjetura que este diseño no puede contrastar. Tampoco todo vínculo con el compromiso parece generalizable: en universitarios peruanos de universidades privadas no se halló relación entre la regulación emocional y el compromiso académico (Marin et al., 2025), lo que indica que la vía hacia el compromiso quizá no se extienda a cualquier forma de regulación y dependa de que esta sea específica del dominio de estudio, como la aquí medida. Por último, la distinción entre la ansiedad ante el aprendizaje y la ansiedad ante la evaluación, que se relacionan de forma diferente con la capacidad de afrontar los contratiempos en matemáticas (Sun y Bai, 2026), sugiere que un indicador único de ansiedad puede ocultar trayectorias distintas para cada forma de ansiedad.

Los resultados coinciden, en cambio, con la correlación negativa entre ansiedad matemática y compromiso conductual descrita con estudiantes de Japón (Shimizu, 2025), y con la observación de que ansiedad y compromiso coexisten en una misma población, con una ansiedad concentrada en las situaciones de evaluación y un nivel alto de compromiso en la educación media superior (Teheran et al., 2024).

La magnitud de las asociaciones exige cautela. La correlación entre autorregulación y ansiedad ($r = -.592$) supera con holgura la que la literatura metaanalítica sitúa entre ansiedad y desempeño matemático ($r = -.28$). La diferencia es esperable, porque aquí ambas variables son percepciones

del propio estudiante medidas con la misma escala de respuesta, mientras que el desempeño suele evaluarse con pruebas externas; por esa razón no conviene comparar directamente ambos órdenes de magnitud, y el modelo debería contrastarse con indicadores objetivos de rendimiento. La invarianza entre mujeres y hombres permite comparar ambos grupos con las mismas escalas, y la estabilidad del efecto directo ante jackknife, submuestreos y remuestreo sugiere que el resultado no depende de casos particulares. La equivalencia de medida transnacional de un inventario de autorregulación entre universitarios de Ecuador, Perú y Venezuela (Regatto et al., 2025) apunta en la misma dirección para extender estas escalas a otros contextos, aunque aquí la invarianza se examinó solo entre sexos. Cabe añadir que los diarios de audio de tres estudiantes de primer año con grados distintos de éxito (Liebendörfer et al., 2023) recuerdan que un modelo centrado en variables promedia trayectorias individuales de regulación que pueden diferir mucho entre estudiantes.

La aportación del estudio tiene alcance general, porque el ingreso a la universidad es un periodo crítico para la ansiedad matemática en cualquier sistema universitario. El modelo identifica dos puntos de intervención complementarios: el entrenamiento en autorregulación, aunque los programas de este tipo se implementan todavía de forma escasa en Sudamérica y con un bajo reporte del tamaño del efecto (Inzunza y Sáez, 2024), y el diseño de experiencias que sostengan el compromiso con los cursos de matemáticas. De ahí se derivan tres implicaciones prácticas:

a) Incorporar en los cursos de primer año sesiones explícitas de planificación y automonitoreo durante las primeras semanas del semestre;

b) Diseñar tareas de dificultad graduada que ofrezcan experiencias tempranas de dominio y favorezcan la implicación sostenida; y

c) Utilizar indicadores de compromiso —asistencia, participación y entrega de tareas— como señales tempranas para derivar a los estudiantes con mayor riesgo de ansiedad a servicios de apoyo.

Estas implicaciones son consistentes con la síntesis de las intervenciones sobre la ansiedad matemática, cuyos tamaños del efecto medios fueron moderados tanto para reducir la ansiedad como para mejorar el rendimiento (Sammallahti et al., 2023).

El estudio presenta limitaciones que acotan sus conclusiones. El diseño transversal impide establecer el orden temporal de la cadena mediacional, y las relaciones recíprocas descritas entre control percibido y ansiedad (Putwain y Wood, 2023) exigen contrastar el modelo con medidas repetidas. El autoinforme de las tres variables amenaza con inflar sus asociaciones, y el muestreo por conveniencia restringe la representatividad de los resultados a los estudiantes analizados, sin generalizarse a la población universitaria del país. Finalmente, el criterio de Fornell y Larcker resultó insuficiente para el aprendizaje autorregulado. Los estudios futuros deberían emplear medidas más extensas de cada dimensión para discriminar la regulación del estudio de la implicación con la disciplina, y revisar las dimensiones del compromiso —vigor, dedicación y absorción— más frecuentes en la literatura (Padilla y García, 2024).

10.6 CONCLUSIONES

En estudiantes universitarios de primer año, el aprendizaje autorregulado se asocia con una menor ansiedad matemática de manera directa y, en una tercera parte de su efecto total, a través del compromiso académico con las matemáticas. El modelo de tres constructos mostró un ajuste excelente y superó con claridad a la alternativa de un solo factor, y el patrón de rutas se mantuvo en los análisis de robustez y ante perturbaciones de la muestra. Este resultado aporta evidencia a favor de la secuencia propuesta por la teoría del control-valor en la transición a la universidad: los recursos de regulación del estudio se vinculan con una mayor implicación en la disciplina y esta, a su vez, con una menor ansiedad, aunque buena parte del efecto de la autorregulación opera de forma directa.

Desde el punto de vista aplicado, los hallazgos sugieren que las intervenciones de acompañamiento académico en matemáticas son más

prometedoras cuando combinan el entrenamiento en autorregulación con el fomento del compromiso con la disciplina, y que los indicadores de implicación pueden servir como señales tempranas de riesgo. Estas conclusiones deben leerse con la prudencia que imponen un diseño transversal, medidas de autoinforme y una muestra por conveniencia, así como la proximidad conceptual entre autorregulación y compromiso.

Quedan abiertas tres preguntas para la investigación futura: si el efecto indirecto se mantiene en diseños longitudinales que permitan establecer el orden temporal y las posibles relaciones recíprocas; qué dimensión del compromiso —conductual, cognitiva o emocional— concentra la mediación; y si el patrón se reproduce con indicadores objetivos de rendimiento, en estudiantes de carreras sin base cuantitativa y en otros sistemas universitarios.

10.7 REFERENCIAS BIBLIOGRÁFICAS

Arevalo, D., Luciano, R., Padilla, C., & Varas, O. (2026). Validación psicométrica de una escala de compromiso académico en estudiantes universitarios de Ecuador y Perú. *Formación universitaria*, *19*(1), 53-64. https://doi.org/10.4067/s0718-50062026000100053

Barroso, C., Ganley, C., McGraw, A., Geer, E., Hart, S., & Daucourt, M. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134-168. https://doi.org/10.1037/bul0000307

Bergqvist, E. (2026). Generalized mathematics self-efficacy and student engagement: the role of self-concept and mathematics anxiety. *Frontiers in Psychology*, *17*, 1852588. https://doi.org/10.3389/fpsyg.2026.1852588

Cahyawati, D., Delima, N., & Gunarto, M. (2023). The impact of undergraduate students' mathematics anxiety and self-concept on their self-regulated learning and academic achievement. *Jurnal Elemen, 9*(1), 183-196. https://doi.org/10.29408/jel.v9i1.6898

Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. https://doi.org/10.1207/s15328007sem0902_5

Dai, W., Li, Z., & Jia, N. (2022). Self-regulated learning, online mathematics learning engagement, and perceived academic control among Chinese junior high school students during the COVID-19 pandemic: A latent profile analysis and mediation analysis. *Frontiers in Psychology, 13,* 1042843. https://doi.org/10.3389/fpsyg.2022.1042843

Daker, R., Gattas, S., Sokolowski, H., Green, A., & Lyons, I. (2021). First-year students' math anxiety predicts STEM avoidance and underperformance throughout university, independently of math ability. *npj Science of Learning, 6*(1), 17. https://doi.org/10.1038/s41539-021-00095-7

Estrada, E., & Paricahua, J. (2023). Compromiso académico en los estudiantes universitarios de la Amazonía Peruana durante el contexto pospandemia. *Revista Digital de Investigación en Docencia Universitaria, 17*(1), e1664. https://doi.org/10.19083/ridu.2023.1664

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement

Error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2022). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology, 114*(2), 346-367. https://doi.org/10.1037/edu0000671

Gamarra, J., Charry, H., Luque, D., & Nina, J. (2026). Entornos virtuales de aprendizaje y aprendizaje autorregulado: percepciones de estudiantes de una universidad pública de Lima, Perú. *Formación universitaria, 19*(4), 33-42. https://doi.org/10.4067/s0718-50062026000400033

Gálvez, F., Pinochet, P., Lepe, N., & Cabrera, H. (2024). Experiencia de aprendizaje autorregulado y su efecto en el rendimiento académico de universitarios de primer año. *Revista Ecuatoriana de Neurologia, 33*(1), 47-54. https://doi.org/10.46997/revecuatneurol33100047

Harahap, L., Andayani, S., & Ekwan, D. (2025). How does self-regulated learning affect students' mathematics anxiety? *Pedagogical Research, 10*(1), em0230. https://doi.org/10.29333/pr/15648

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Inzunza, B., & Sáez, F. (2024). Programas de entrenamiento para promover la autorregulación del aprendizaje en estudiantes universitarios: una revisión sistemática. *Revista Iberoamericana de Educación Superior*, 186-211. https://doi.org/10.22201/iisue.20072872e.2024.44.1898

Klee, H., Buehl, M., & Miller, A. (2022). Strategies for alleviating students' math anxiety: Control-value theory in practice. *Theory Into Practice*, *61*(1), 49-61. https://doi.org/10.1080/00405841.2021.1932157

Liebendörfer, M., Kempen, L., & Schukajlow, S. (2023). First-year university students' self-regulated learning during the Covid-19 pandemic: a qualitative longitudinal study. *ZDM – Mathematics Education*, *55*(1), 119-131. https://doi.org/10.1007/s11858-022-01444-5

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Marin, D., Alvarez, M., & Araujo, E. (2025). Cansancio y regulación emocional como predictores del compromiso académico en estudiantes universitarios. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, *15*(30). https://doi.org/10.23913/ride.v15i30.2264

Mucha, L., Surichaqui, M., Guerra, W., Aragón, A., & Huamán, M. (2025). Competencias de Aprendizaje Autorregulado en Estudiantes Universitarios: Análisis Factorial y de Comunalidad. *Estudios pedagógicos (Valdivia)*, *51*(3), 407-425. https://doi.org/10.4067/s0718-07052025000300407

Núñez, M., & Guilera, G. (2023). Development and Validation of the Brief Math Anxiety Scale (BMAS) in University Students. *Psicothema*, *35*(4), 406-413. https://doi.org/10.7334/psicothema2022.434

Padilla, J., & García, J. (2024). Exploración del engagement académico de estudiantes: una revisión sistemática. *Revista Educación*, *48*(2), 1-20. https://doi.org/10.15517/revedu.v48i2.58455

Pekrun, R. (2024). Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions. *Educational Psychology Review*, *36*(3), 83. https://doi.org/10.1007/s10648-024-09909-7

Pelegrina, S., Toffalini, E., & Martín, M. (2026). Stability of math anxiety and its relation with math performance over time: A meta-analysis of longitudinal studies. *Journal of Educational Psychology*, *118*(5), 677-702. https://doi.org/10.1037/edu0001008

Putwain, D., & Wood, P. (2023). Anxiety in the mathematics classroom: reciprocal relations with control and value, and relations with subsequent achievement. *ZDM – Mathematics Education*, *55*(2), 285-298. https://doi.org/10.1007/s11858-022-01390-2

Quiroz, J., Vilca, L., & Travezaño, A. (2026). The Abbreviated Math Anxiety Scale (AMAS): New psychometric evidence in Peruvian adolescents. *Suma Psicológica*, *33*(1), 26-36. https://doi.org/10.14349/sumapsi.2026.v33.n1.3

Regatto, J., Viteri, V., & Moreta, R. (2025). Equivalencia de medida transnacional del Inventario de Procesos de Autorregulación del Aprendizaje en universitarios de Ecuador, Perú y Venezuela.

Psychology, Society & Education, *17*(2), 66-75. https://doi.org/10.21071/pse.v17i2.17442

Rocha, J., Mamani, A., & Damian, E. (2023). Validación del cuestionario de aprendizaje autorregulado [CAA] en estudiantes universitarios. *Revista peruana de investigación e innovación educativa*, *3*(2), e25298. https://doi.org/10.15381/rpiiedu.v3i2.25298

Sammallahti, E., Finell, J., Jonsson, B., & Korhonen, J. (2023). A meta-analysis of math anxiety interventions. *Journal of Numerical Cognition*, *9*(2), 346-362. https://doi.org/10.5964/jnc.8401

Shi, F., & Li, J. (2026). Self-regulated learning and academic engagement among university students: a latent mediation model of achievement emotions. *Frontiers in Psychology*, *17*, 1822649. https://doi.org/10.3389/fpsyg.2026.1822649

Shimizu, Y. (2025). Relation Between Mathematics Self-Efficacy, Mathematics Anxiety, Behavioural Engagement, and Mathematics Achievement in Japan. *Psychology International*, *7*(2), 36. https://doi.org/10.3390/psycholint7020036

Sun, Y., & Bai, Y. (2026). Distinguishing math learning and evaluation anxiety: a structural equation modeling study on pathways from teacher support to academic buoyancy. *Frontiers in Psychology*, *17*, 1829556. https://doi.org/10.3389/fpsyg.2026.1829556

Szucs, D., & Toffalini, E. (2023). Maths anxiety and subjective perception of control, value and success expectancy in mathematics. *Royal Society Open Science*, *10*(11), 231000. https://doi.org/10.1098/rsos.231000

Teheran, V., Sánchez, J., & Díaz, A. (2024). Ansiedad matemática y engagement académico en estudiantes de educación media superior en México. *Educatio Siglo XXI, 42*(2), 97-120. https://doi.org/10.6018/educatio.591541

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology, 66*, 101976. https://doi.org/10.1016/j.cedpsych.2021.101976

Veiga, F., Wong, Z., Reeve, J., Jimerson, S., Leite, A., Perales, J., Valente, S., & Martínez, I. (2026). Secondary School Students' Engagement in Learning Activities: Validation of a Short Scale. *Education Sciences, 16*(2), 279. https://doi.org/10.3390/educsci16020279

Zamora, S., Cadena, M., Franci, A., & Cibrián, T. (2023). Prevalencia de ansiedad matemática en universitarios. *Nova Scientia, 15*(31), 1-14. https://doi.org/10.21640/ns.v15i31.3185

Zientek, L., & Lane, F. (2026). Revisiting Self-Regulation, Math Self-Efficacy, and Mathematics Anxiety With a Sample of University Students in the United States: A Structural Equation Model of Mathematics Anxiety. *Journal for STEM Education Research*. https://doi.org/10.1007/s41979-026-00182-7

Zimmerman, B. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice, 41*(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

CAPÍTULO XI

EFECTOS DEL MÉTODO SINGAPUR SOBRE LA RESOLUCIÓN DE PROBLEMAS MATEMÁTICOS EN ESTUDIANTES DE PRIMARIA: UN ESTUDIO CUASIEXPERIMENTAL EN AULAS DE PERÚ, 2025

Ibeth Yusely Garcia Puma

Universidad Nacional de San Antonio Abad del Cusco

Resumen

Antecedentes. La resolución de problemas verbales es el punto en que la matemática de primaria con mayor frecuencia se quiebra, y la instrucción explícita basada en representaciones puede mejorarla. El método Singapur combina la secuencia concreto-pictórico-abstracta con el modelo de barras, pero la evidencia a escala de aula que distinga sus efectos por fase y atienda el anidamiento de estudiantes en aulas es escasa fuera de Asia oriental. **Objetivo.** Estimar el efecto de un programa de 12 semanas del método Singapur sobre la resolución de problemas de estudiantes de cuarto grado, en total y por fase (comprensión, planificación, ejecución y revisión). **Método.** Diseño cuasiexperimental pretest-postest con grupo de comparación no equivalente en nueve escuelas públicas peruanas durante 2025 (18 aulas; 449 estudiantes analizados). El análisis principal fue un análisis de covarianza de dos niveles ajustado por el pretest, complementado con modelos por fase en secuencia fija, ocho análisis de robustez y una estimación sobre el dominio. **Resultados.** Con ajuste por pretest y conglomeración, el método Singapur superó a la comparación en 5.93 puntos (IC 95 % [2.23, 9.64]; $g = 0.40$). Los efectos fueron mayores en planificación ($g = 0.47$) y comprensión ($g = 0.31$), menores en ejecución ($g = 0.28$) y nulos en revisión ($g = 0.14$). La razón de momios ajustada de alcanzar el dominio fue 2.52. **Conclusiones.** El método mejora la resolución de problemas sobre todo en las fases representacionales; la fase reflexiva requiere instrucción propia.

Palabras clave: matemática Singapur; modelo de barras; concreto-pictórico-abstracto; resolución de problemas verbales; educación primaria; análisis multinivel.

Abstract

Background. *Word-problem solving is where primary mathematics most often breaks down, and explicit representation-based instruction can improve it. The Singapore method combines a concrete–pictorial–abstract sequence with*

the bar model, yet classroom-scale evidence that separates its effects by problem-solving phase while accounting for the nesting of students in classrooms is scarce outside East Asia. **Objective**. *To estimate the effect of a 12-week Singapore-method programme on fourth-grade students' problem solving, overall and by phase (understanding, planning, execution and looking back).* **Method**. *Quasi-experimental pretest–posttest design with a non-equivalent comparison group in nine Peruvian public schools during 2025 (18 classrooms; 449 students analysed). The primary analysis was a two-level analysis of covariance adjusted for the pretest, complemented by phase-specific models in a fixed sequence, eight robustness analyses and an estimate of proficiency.* **Results**. *Adjusting for the pretest and clustering, the Singapore condition scored 5.93 points higher (95 % CI [2.23, 9.64]; Hedges' g = 0.40). Effects were largest for planning (g = 0.47) and understanding (g = 0.31), smaller for execution (g = 0.28) and absent for looking back (g = 0.14). The adjusted odds ratio of reaching proficiency was 2.52.* **Conclusions.** *The method improves problem solving mainly through the representational phases; the reflective phase requires instruction of its own.*

Keywords: *Singapore mathematics; bar model; concrete–pictorial–abstract; word-problem solving; primary education; multilevel analysis*

11.1 INTRODUCCIÓN

11.1.1 Planteamiento del problema

Resolver un problema verbal exige leer una situación, construir una representación de las cantidades y relaciones que describe, seleccionar operaciones y verificar que la respuesta sea sensata. Cada uno de estos pasos moviliza recursos distintos. Las características lingüísticas y numéricas de la tarea modifican su dificultad (Vessonen et al., 2024); las características cognitivas y académicas individuales se asocian con el desempeño en problemas verbales (Vessonen et al., 2025b); y el conocimiento de números naturales y de fracciones conforma una jerarquía que conduce a la resolución de problemas (Xu et al., 2024). A ello se suman las funciones ejecutivas: la inhibición y la flexibilidad cognitiva contribuyen al

desempeño en la resolución de problemas en la escuela primaria (Schäfer et al., 2024). La resolución de problemas es, por tanto, el lugar donde el conocimiento matemático debe orquestarse y no solo recuperarse, y es allí donde fracasa una proporción elevada de estudiantes.

La dificultad es especialmente visible en los sistemas educativos con resultados inferiores a los referentes internacionales. Perú permanece por debajo del promedio de la OCDE en matemáticas en el ciclo más reciente de PISA (OECD, 2023), y en la región se han examinado, en un estudio exploratorio en Chile, las actitudes de estudiantes de primaria superior hacia la resolución de problemas matemáticos (Saadati et al., 2025). La pregunta pertinente para tales sistemas no es si la resolución de problemas puede enseñarse, sino qué enfoque instruccional la modifica a escala de aula y en qué fases del proceso de solución se produce el cambio.

11.1.2 Antecedentes y vacío de conocimiento

El método Singapur articula en un currículo coherente varios componentes de efecto plausible. Su secuencia concreto-pictórico-abstracta transita de los manipulativos a los dibujos esquemáticos antes del trabajo simbólico, y el método del modelo, o modelo de barras, representa relaciones parte-todo y de comparación mediante rectángulos cuya longitud codifica las cantidades (Ng, 2022). El dibujo esquemático busca apoyar la decodificación de la información y la traducción de las relaciones verbales en operaciones (Baysal y Sevinç, 2022), y trabajos recientes revisan los fundamentos teóricos del principio concreto-pictórico-abstracto (Leong et al., 2026). Los estudios de esta secuencia en entornos con recursos limitados informan mejoras en rendimiento y motivación (Iyamuremye y Burns, 2025), y se han examinado los efectos de manipulativos concretos y virtuales en grupos de bajo, medio y alto rendimiento (Ahmad y Siller, 2024). La instrucción en diagramas adecuados al problema, por su parte, atiende el hecho de que los estudiantes rara vez construyen diagramas por iniciativa propia (Ayabe et al., 2022).

La propia representación resulta exigente: las representaciones con modelo de barras de docentes en formación para problemas de fracciones se han convertido en objeto de estudio (Lizano y Sevinç, 2025), y las revisiones curriculares muestran que componentes como la formulación de problemas reciben atención limitada en el currículo de Singapur (Toh y Chua, 2025). En América Latina, una revisión de la literatura sobre el modelo de barras atribuye a este contribuciones a la construcción del conocimiento, a la profundización de conceptos y relaciones matemáticas y a las habilidades de resolución de problemas (Richit y Richit, 2022).

De estos antecedentes se desprenden tres vacíos.

1) Primero, la mayoría de las evaluaciones informa una única puntuación total, de modo que se desconoce si el método actúa sobre las fases representacionales a las que apunta (comprensión y planificación) o de manera difusa a lo largo del ciclo de solución.

2) Segundo, las implementaciones en aula asignan grupos intactos, y los análisis que ignoran la conglomeración sobrestiman la precisión.

3) Tercero, se carece de estimaciones de aula, sensibles a los conglomerados y procedentes de sistemas ajenos a Asia oriental, que permitan calibrar las expectativas.

11.1.3 Pregunta, objetivos y contribución

El estudio plantea la pregunta siguiente: ¿cuál es el efecto del método Singapur sobre la resolución de problemas de estudiantes de cuarto grado y en qué fases del proceso de resolución se produce? Los objetivos son:

a) Estimar el efecto sobre la puntuación total de resolución de problemas, ajustando por el rendimiento inicial y por la conglomeración en aulas;

b) Estimar los efectos específicos de cada fase, siguiendo el ciclo de comprender, planificar, ejecutar y revisar (Pólya, 1945); y

c) Expresar el efecto como cambio en la proporción de estudiantes que alcanzan el dominio.

La contribución es una estimación resuelta por fase y sensible a los conglomerados, que vincula el mecanismo atribuido al modelo de barras con la fase en la que se observa el cambio, junto con una verificación de los parámetros de diseño que futuras evaluaciones en aula podrán reutilizar.

11.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

11.2.1 La resolución de problemas como ciclo de cuatro fases

Siguiendo a Pólya (1945), la resolución de problemas se concibe como un ciclo de comprender el problema, concebir un plan, ejecutarlo y mirar hacia atrás. Las fases descansan en competencias diferentes. Comprender y planificar dependen de construir una representación de la situación y de las relaciones entre cantidades, proceso en el que intervienen las demandas lingüísticas de la tarea y el conocimiento numérico previo, y que exige inhibir información irrelevante y alternar entre estrategias. La ejecución depende de la fluidez computacional, y la revisión, de la supervisión y evaluación de la solución obtenida. La medición específica por fase permite someter a prueba el mecanismo de una intervención en lugar de darlo por supuesto.

11.2.2 Cómo actúa el método Singapur sobre el ciclo

El modelo de barras externaliza la estructura del problema: el estudiante identifica la incógnita, las cantidades conocidas y su relación, y las dibuja antes de elegir una operación (Ng, 2022). Esto convierte a las fases representacionales en su punto natural de acción. La secuencia concreto-pictórico-abstracta sostiene esa traducción al pasar de la manipulación al esquema y de este al símbolo (Leong et al., 2026), y la evidencia sobre el modelo de barras indica que reduce los errores en problemas verbales (Baysal y Sevinç, 2022). En la misma línea, enseñar a reconocer y dibujar la estructura del problema mejora el desempeño cuando el diagrama es

adecuado a la tarea (Ayabe et al., 2022). Se espera que la ejecución se beneficie de manera indirecta, porque un plan correcto reduce los errores procedimentales, mientras que la revisión no es blanco directo de la representación y requeriría instrucción explícita de verificación. De estas consideraciones derivan dos hipótesis:

- **H1.** Los estudiantes que reciben el método Singapur obtienen puntuaciones de resolución de problemas en el postest superiores a las de quienes reciben la instrucción habitual, con ajuste por el pretest.
- **H2.** El efecto es mayor en las fases de comprensión y planificación que en las de ejecución y revisión.

El análisis del dominio expresa el efecto de H1 en una escala referida a criterio; se informa como estimación con su intervalo de confianza y no como contraste de una hipótesis adicional.

11.3 MATERIALES Y MÉTODOS

11.3.2 Diseño y participantes

Se empleó un diseño cuasiexperimental pretest-postest con grupo de comparación no equivalente y dos niveles de anidamiento (estudiantes en aulas). En cada una de nueve escuelas primarias públicas urbanas de Perú participaron, durante 2025, dos aulas paralelas de cuarto grado: una implementó el método Singapur y la otra continuó con la instrucción habitual basada en el currículo nacional. La condición se asignó por aula según el horario escolar, no al azar, por lo que se anticiparon diferencias iniciales y se ajustaron estadísticamente. La selección de escuelas fue de conveniencia, por acuerdo con las direcciones, y no se infiere a una población más amplia.

El tamaño muestral se calculó a priori para un análisis de covarianza de dos niveles con asignación por aula. El efecto mínimo de interés fue d = 0.40, fijado por debajo de los valores meta-analíticos porque los efectos de campo suelen ser menores (Kraft, 2020). Para una comparación individual

se requieren 98.1 estudiantes por condición ($\alpha = .05$ bilateral; $1 - \beta = .80$); este valor se multiplicó por el efecto de diseño $1 + (m - 1)\rho$ y por $1 - r^2$, donde r es la correlación de las medias de aula entre pretest y postest (Teerenstra et al., 2012). Con $m = 25$, $\rho = .15$, en el rango compilado para logros educativos (Hedges y Hedberg, 2007), una correlación pretest-postest de .60 entre estudiantes y .80 entre aulas, el factor de varianza fue 1.922 y el requisito, 7.54 aulas por condición. Recalculada la potencia con la distribución t no central con $2k - 3$ grados de libertad, ocho aulas por condición dieron $1 - \beta = .760$ y nueve, .816, de modo que se fijaron nueve aulas por condición y 450 estudiantes. La muestra analizada fue de 449 (227 método Singapur; 222 comparación) de 466 matriculados, con 23 a 28 estudiantes por aula (media 24.94). Las pruebas por fase se plantearon en secuencia fija, y el dominio se informó solo como estimación, porque su contraste separado carecería de potencia suficiente; la moderación por sexo no se examinó por el mismo motivo.

11.3.3 Instrumento

La resolución de problemas se midió con dos formas paralelas (pretest y postest) de una prueba de ocho problemas verbales no rutinarios alineados con el currículo de cuarto grado. Una rúbrica analítica calificó cada fase de 0 a 2: comprensión, planificación, ejecución y revisión. Las fases puntúan de 0 a 16 y el total, de 0 a 64. La consistencia interna se estimó con α de Cronbach y ω de McDonald a partir de un modelo unifactorial por fase, cuyo ajuste se inspeccionó porque ω depende de un modelo correctamente especificado (Bell et al., 2024); el total se evaluó con α estratificado y la concordancia entre evaluadores, con el coeficiente de correlación intraclase en una submuestra aleatoria de 76 guiones calificados por un segundo evaluador. El dominio se definió a priori como un total de al menos 39 puntos (60 % del máximo).

11.3.4 Intervención y procedimiento

El programa constó de 36 sesiones de 90 minutos (tres semanales durante 12 semanas) a cargo de los docentes regulares, tras 16 horas de

formación. Cada unidad siguió la secuencia concreto-pictórico-abstracta, con modelos de barras de parte-todo y comparación y una rutina explícita de cuatro pasos al cierre de cada lección; la implementación se observó con una lista de cotejo en una quinta parte de las sesiones. Las aulas de comparación dedicaron el mismo tiempo a resolver problemas con el texto escolar, sin modelos de barras. El pretest se aplicó la semana previa y el postest la posterior, por evaluadores externos, y la calificación fue ciega a la condición.

11.3.5 Estrategia analítica

La equivalencia inicial se evaluó con la g de Hedges y la h de Cohen; diferencias entre $|0.05|$ y $|0.25|$ se trataron como ajustables. El análisis principal fue un ANCOVA de dos niveles con intercepto aleatorio por aula, pretest centrado en la media general y condición como predictor de aula, estimado por máxima verosimilitud restringida (McNeish, 2026), con la prueba de la condición sobre $J - 3 = 15$ grados de libertad (Yamamoto y Miyazaki, 2025). El tamaño del efecto fue la g de Hedges con intervalo por bootstrap de conglomerados ($B = 1000$), interpretada conforme a Hedges (2025). Los efectos por fase se estimaron con el mismo modelo y se contrastaron en la secuencia fija planificación, comprensión, ejecución y revisión, cada uno con $\alpha = .05$ y solo si todos los anteriores, comenzando por H1, eran significativos. Se examinaron los supuestos mediante asimetría, curtosis, Shapiro-Wilk, Levene, interacción condición × pretest y término cuadrático. La robustez se evaluó con ocho especificaciones: ANCOVA de un nivel con errores estándar basados en el modelo y robustos a conglomerados, bootstrap-t de conglomerados salvaje, ANCOVA por rangos de Quade, estimación M de Huber, modelo de dos niveles con covariables, análisis sobre las 18 medias de aula y modelo de dos niveles de puntuaciones de cambio. El dominio se analizó con ecuaciones de estimación generalizadas logísticas de correlación intercambiable. Por último, se compararon los parámetros de diseño observados con los planificados. Los análisis se realizaron en Python.

11.4 RESULTADOS

11.4.1 Muestra, equivalencia inicial y atrición

De los 466 matriculados, 17 (3.65 %) no completaron el postest, 11 en el grupo Singapur y 6 en el de comparación, una atrición diferencial de 1.99 puntos porcentuales. Los estudiantes perdidos tenían puntuaciones de pretest menores que los retenidos (23.18 frente a 27.88; $g = -0.31$). La Tabla 11-1 compara los grupos analizados. Edad, sexo e índice socioeconómico resultaron equilibrados, mientras que las aulas de comparación iniciaron con puntuaciones mayores en el total y en todas las fases, con diferencias dentro del rango que exige ajuste estadístico.

Tabla 11-1

Características iniciales de los estudiantes analizados según condición

Característica		Método Singapur (n = 227)	Comparación (n = 222)	Tamaño del efecto	p
Edad (años), *M* (*DE*)		9.59 (0.36)	9.60 (0.36)	$g = -0.03$	.759
Niñas, *n* (%)		112 (49.3)	103 (46.4)	$h = 0.06$	.533
Índice socioeconómico (*z*), *M* (*DE*)		−0.11 (0.94)	−0.14 (0.87)	$g = 0.04$	.673
Pretest *M* (*DE*)	total (0–64)	26.37 (15.66)	29.41 (14.95)	$g = -0.20$	.036
	comprensión (0–16)	9.20 (4.75)	9.91 (4.45)	$g = -0.15$	.106
	planificación (0–16)	6.92 (4.87)	7.57 (4.66)	$g = -0.14$	.151
	ejecución (0–16)	6.22 (4.71)	7.02 (4.36)	$g = -0.18$	.062
	revisión (0–16),	4.04 (3.96)	4.92 (4.13)	$g = -0.22$	.021

Nota: Pruebas t de Welch para variables continuas y χ^2 para sexo. g = g de Hedges (Singapur menos comparación); h = h de Cohen.

11.4.2 Propiedades de medición

Según la Tabla 11-2, cada subescala de fase fue internamente consistente en ambas ocasiones (α y ω entre .885 y .912), con cargas factoriales estandarizadas entre 0.62 y 0.79 y correlaciones ítem-resto entre 0.58 y 0.74. Los modelos unifactoriales mostraron CFI ≥ .987 y RMSEA ≤ .049. El α estratificado del total fue .966 en el pretest y .965 en el postest, y la concordancia entre evaluadores en los 76 guiones fue alta, con una concordancia exacta a nivel de ítem de 93.3 %.

Tabla 11-2

Fiabilidad y concordancia entre evaluadores de la prueba de resolución de problemas

Fase	pre		Cargas pre (mín–máx)	post		*r* ítem–resto post (mín–máx)	CFI / RMSEA post	CCI entre evaluadores [IC 95 %]
	α	ω		α	ω			
Comprensión	.903	.904	0.72–0.77	.895	.898	0.63–0.71	.987 / .049	.983 [.97, .99]
Planificación	.910	.911	0.69–0.79	.904	.904	0.65–0.72	.994 / .034	.987 [.98, .99]
Ejecución	.896	.897	0.68–0.76	.905	.905	0.67–0.72	.997 / .026	.986 [.98, .99]
Revisión	.885	.886	0.62–0.74	.911	.912	0.68–0.74	.997 / .027	.978 [.96, .99]
Total (α estratificado)	.966	—	—	.965	—	—	—	.994 [.99, 1.00]

Nota: ω a partir de un modelo unifactorial por fase (ocho problemas). CCI = efectos aleatorios bidireccionales, acuerdo absoluto, medida única, sobre 76 guiones de postest calificados por duplicado. Las correlaciones ítem-resto del pretest fluctuaron entre 0.58 y 0.74 y el CFI del pretest, entre .992 y 1.000.

11.4.3 Estadísticos descriptivos

La Tabla 11-3 presenta las puntuaciones por condición y ocasión. Ambos grupos mejoraron, pero las aulas del método Singapur superaron a las de comparación en el postest pese a su inicio más bajo. Las ganancias fueron mayores en planificación y comprensión y menores en revisión en ambos grupos, y la correlación pretest-postest fue de .538 para el total y de entre .301 y .417 para las fases. La Figura 11-1 (Panel A) ilustra estos cambios.

Tabla 11-3

Puntuaciones de resolución de problemas por condición y ocasión

Resultado	Singapur *M* (*DE*)			Comparación *M* (*DE*)			r pre–post
	pre	post [IC 95 %]	ganancia	pre	post [IC 95 %]	ganancia	
Total (0–64)	26.37 (15.66)	37.84 (15.12) [35.87, 39.80]	11.46 (14.06)	29.41 (14.95)	33.46 (14.86) [31.51, 35.42]	4.05 (14.34)	.538
Comprensión (0–16)	9.20 (4.75)	12.18 (3.98) [11.66, 12.70]	2.98 (4.94)	9.91 (4.45)	11.10 (4.33) [10.53, 11.67]	1.19 (5.10)	.332
Planificación (0–16)	6.92 (4.87)	10.78 (4.66) [10.17, 11.38]	3.85 (4.98)	7.57 (4.66)	8.88 (4.37) [8.31, 9.46]	1.32 (5.04)	.395
Ejecución (0–16)	6.22 (4.71)	8.99 (4.78) [8.37, 9.61]	2.77 (4.98)	7.02 (4.36)	7.99 (4.68) [7.37, 8.60]	0.97 (4.92)	.417
Revisión (0–16)	4.04 (3.96)	5.89 (4.54) [5.30, 6.48]	1.86 (4.95)	4.92 (4.13)	5.50 (4.73) [4.87, 6.12]	0.57 (5.30)	.301

Nota: IC 95 % de la media del postest (aproximación normal, nivel de estudiante).

11.4.4 Verificación de supuestos

La Tabla 11-4 resume los diagnósticos. Las pendientes de regresión fueron homogéneas en todos los resultados y el término cuadrático del pretest fue despreciable, lo que respalda el ANCOVA lineal. Con 449 estudiantes, Shapiro-Wilk detectó desviaciones menores de la normalidad, con asimetría y curtosis dentro de ±0.83. Comprensión mostró asimetría negativa (–0.81), compatible con un techo en el postest, y varianzas residuales desiguales (Levene $p = .035$).

Tabla 11-4

Diagnósticos de los supuestos del análisis de covarianza

Resultado	Asimetría	Curtosis	Shapiro–Wilk W (p)	Levene F (p)	Condición × pretest F (p)	Pretest cuadrático t (p)	FIV
Total	−0.14	−0.46	0.993 (.029)	0.13 (.716)	0.14 (.710)	−0.57 (.569)	1.010
Comprensión	−0.81	−0.01	0.940 (< .001)	4.46 (.035)	0.00 (.955)	1.26 (.207)	1.006
Planificación	−0.36	−0.56	0.980 (< .001)	0.16 (.687)	0.95 (.331)	−0.40 (.690)	1.005
Ejecución	−0.10	−0.83	0.983 (< .001)	0.04 (.852)	0.04 (.843)	−0.07 (.942)	1.008
Revisión	0.51	−0.55	0.962 (< .001)	0.18 (.669)	0.19 (.665)	−1.51 (.132)	1.012

Nota: Diagnósticos sobre los residuos del ANCOVA de un nivel (postest sobre pretest y condición). Prueba de Levene centrada en la mediana. FIV = factor de inflación de la varianza entre condición y pretest

11.4.5 Efecto sobre la resolución total de problemas (H1)

La Tabla 11-5 presenta el ANCOVA de dos niveles principal. El pretest predijo con fuerza el postest y, con el pretest en su media general, la media ajustada fue de 38.56 en la condición Singapur y de 32.63 en la de comparación. La diferencia ajustada de 5.93 puntos fue estadísticamente significativa y correspondió a $g = 0.40$. Añadir la condición mejoró el modelo, $\chi^2(1) = 9.79$, $p = .002$. Tras el ajuste, las aulas explicaron el 4.9 % de la varianza residual, y el modelo explicó el 30.6 % (marginal) y el 34.0 % (condicional) de la varianza del postest. La Figura 11-2 muestra la geometría del ajuste, la robustez de la estimación y los contrastes dentro de cada escuela.

Tabla 11-5

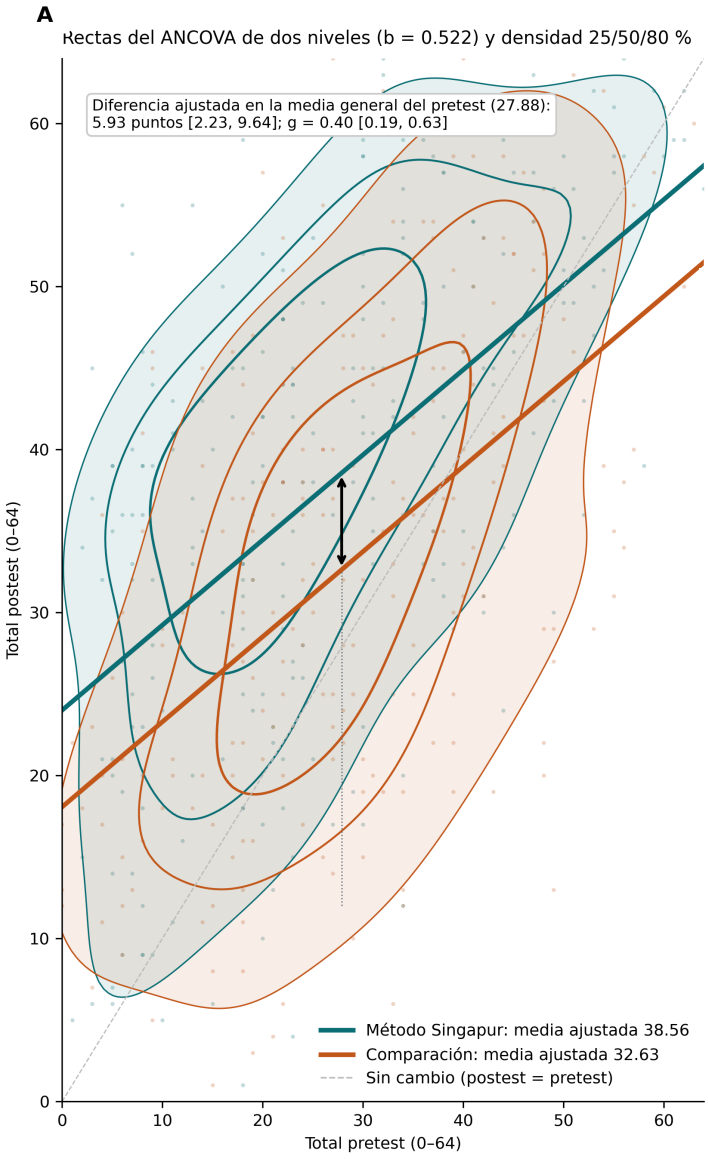
Análisis de covarianza de dos niveles de la puntuación total del postest

Parámetro	Estimación	EE	*t* (*gl*)	p	IC 95 %
Intercepto (media ajustada, comparación)	32.63	1.23	—	—	—
Pretest total (centrado en la media general)	0.522	0.040	12.99	< .001	—
Condición (Singapur vs. comparación)	5.93	1.74	3.41 (15)	.004	[2.23, 9.64]
Media ajustada, Singapur	38.56	—	—	—	—
g de Hedges (bootstrap de conglomerados)	0.40	—	—	—	[0.19, 0.63]
Varianza de aula (τ_{00})	7.61	—	—	—	—
Varianza residual (σ^2)	147.71	—	—	—	—
Correlación intraclase condicional	.049	—	—	—	—
Prueba de razón de verosimilitud de la condición, $\chi^2(1)$	9.79	—	—	.002	—
R^2 marginal / condicional	.306 / .340	—	—	—	—

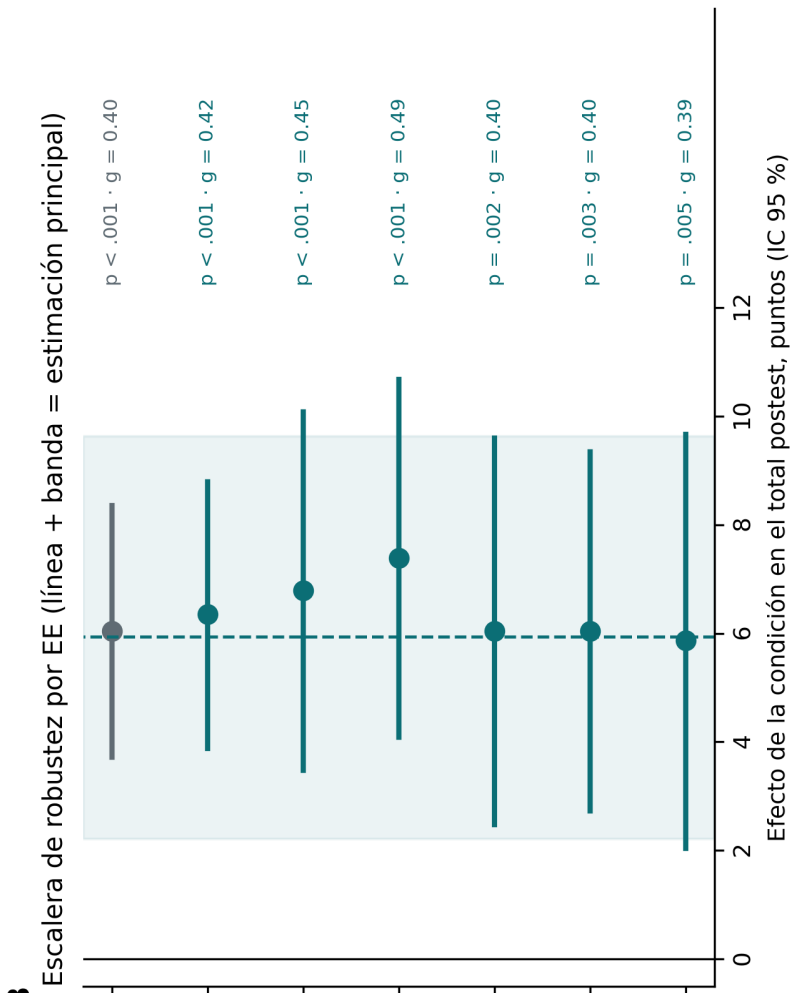
Nota: Máxima verosimilitud restringida; 18 aulas, 449 estudiantes. El efecto de la condición se contrastó con J − 3 = 15 grados de libertad. La g de Hedges usa la desviación típica combinada del postest (14.99); su intervalo proviene de 1000 réplicas bootstrap que remuestrean aulas dentro de cada condición. La razón de verosimilitud compara ajustes por máxima verosimilitud con y sin condición.

Figura 11-1

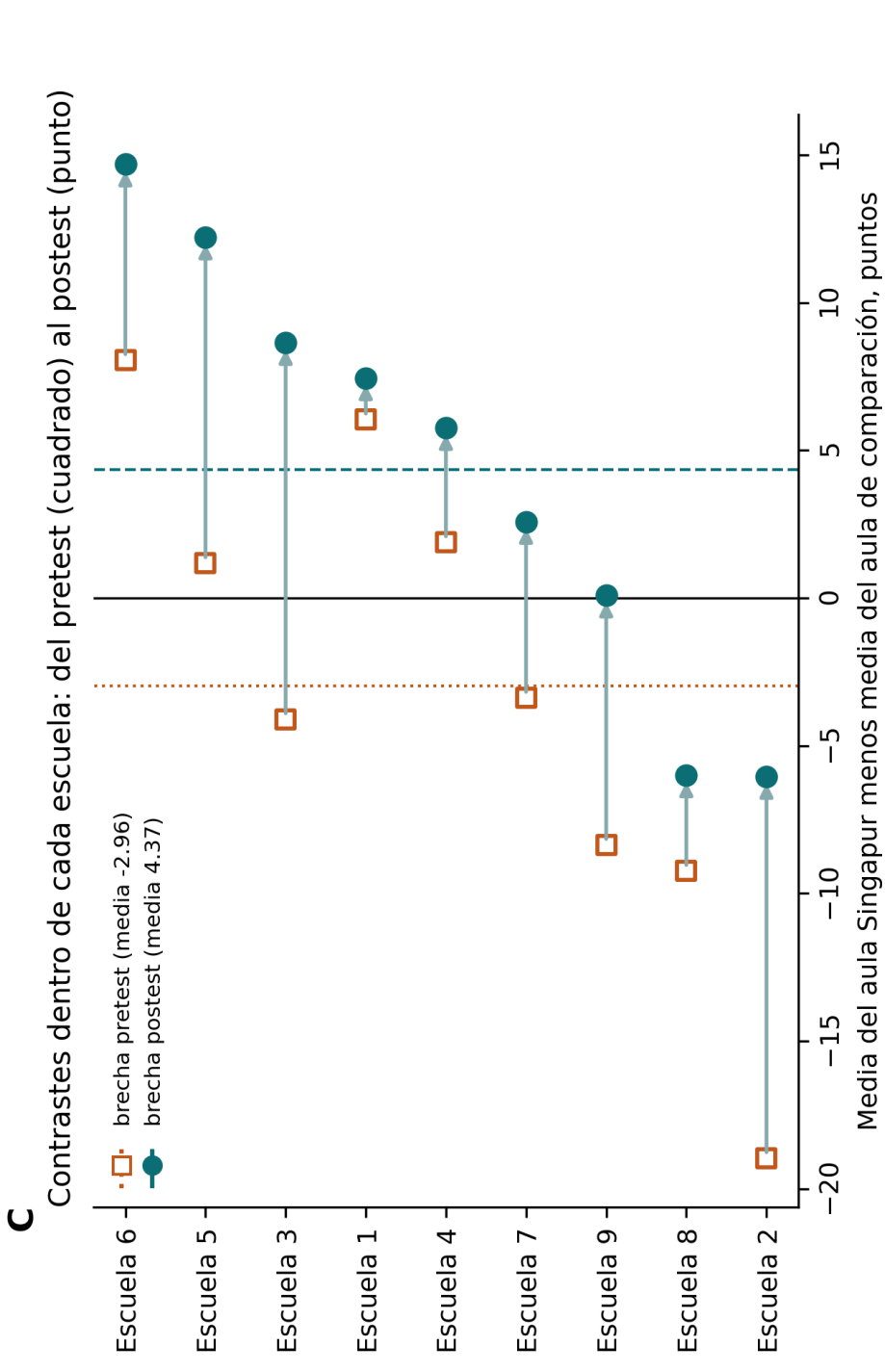
Geometría del ajuste, robustez de la especificación y contrastes dentro de cada escuela para la resolución total de problema



Nota. Análisis de covarianza de dos niveles de la puntuación total del postest sobre la condición y el pretest, con estudiantes anidados en aulas (N = 449 estudiantes, 18 aulas en 9 escuelas). Panel A: dispersión pretest–postest con las regiones de máxima densidad del 25 %, 50 % y 80 % de cada condición y las rectas del análisis de covarianza con pendiente común b = 0.522 (diagonal discontinua: postest igual al pretest); en la media general del pretest (27.88) las medias ajustadas son 38.56 (método Singapur) y 32.63 (comparación), una diferencia ajustada de 5.93 puntos (IC 95 % [2.23, 9.64]; g = 0.40, IC 95 % por bootstrap de conglomerados [0.19, 0.63]). Fuente: Tablas 11-5 y 11-6.



Nota. Análisis de covarianza de dos niveles de la puntuación total del postest sobre la condición y el pretest, con estudiantes anidados en aulas (N = 449 estudiantes, 18 aulas en 9 escuelas). Panel B: efecto de la condición bajo las siete especificaciones, en puntos del postest, ordenadas por error estándar (EE), con IC 95 %; el modelo de un nivel que ignora los conglomerados (gris) muestra el error estándar subestimado, y la línea discontinua y la banda marcan la estimación principal de dos niveles; el análisis por rangos de Quade (p = .002) se informa en la escala de rangos en la Tabla 11-7. Fuente: Tablas 11-5 y 11-6.



Nota. Análisis de covarianza de dos niveles de la puntuación total del postest sobre la condición y el pretest, con estudiantes anidados en aulas (N = 449 estudiantes, 18 aulas en 9 escuelas). Panel *C*: en cada una de las nueve escuelas, diferencia entre la media del aula del método Singapur y la del aula de comparación en el pretest (cuadrado; media -2.96 puntos) y en el postest (punto; media sin ajustar 4.37 puntos). Fuente: Tablas 11-5 y 11-6.

11.4.6 Efectos por fase de resolución de problemas (H2)

Al rechazarse H1, se ingresó a la secuencia fija. Según la Tabla 11-6, planificación mostró el mayor efecto (g = 0.47; p = .002), seguida de comprensión (g = 0.31; p = .010) y ejecución (g = 0.28; p = .045), cada una significativa en su paso, y la secuencia se detuvo en revisión (g = 0.14; p = .358). Ninguna prueba comparó las fases entre sí. La Figura 11-1 reúne los cambios, los efectos ajustados y las distribuciones del postest por fase.

Tabla 11-6

Análisis de covarianza de dos niveles por fase de resolución de problemas

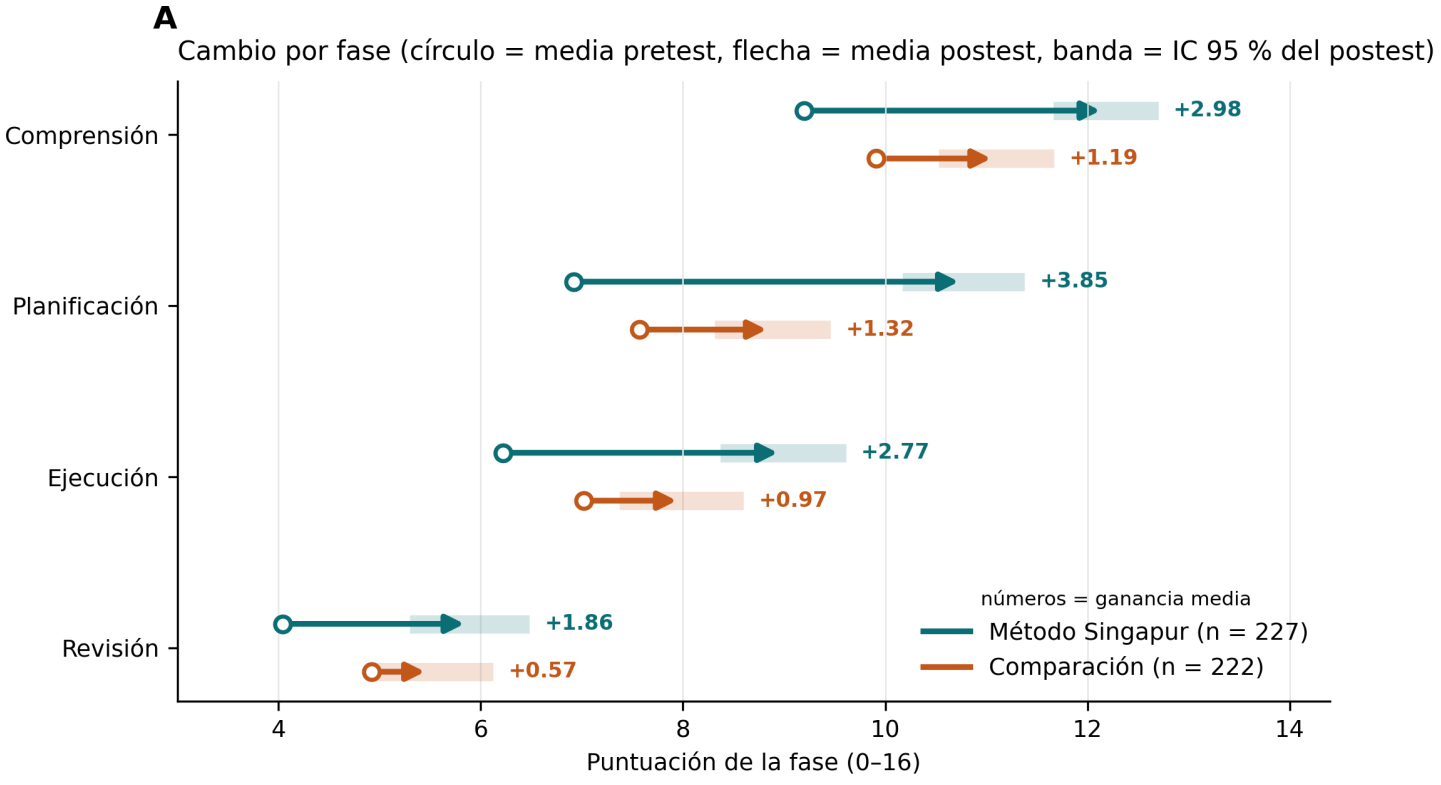
Fase	Media ajustada		Diferencia [IC 95 %]	*EE*	t(15)	*p*
	Singapur	comparación				
Comprensión	12.28	10.99	1.30 [0.36, 2.23]	0.44	2.95	.010
Planificación	10.87	8.75	2.12 [0.93, 3.31]	0.56	3.79	.002
Ejecución	9.14	7.81	1.34 [0.04, 2.64]	0.61	2.19	.045
Revisión	6.02	5.36	0.66 [–0.82, 2.13]	0.69	0.95	.358

Fase	Paso en la secuencia fija	g de Hedges	CCI condicional
Comprensión	2: rechazada	0.31	.017
Planificación	1: rechazada	0.47	.044
Ejecución	3: rechazada	0.28	.053
Revisión	4: no rechazada (la secuencia se detiene)	0.14	.072

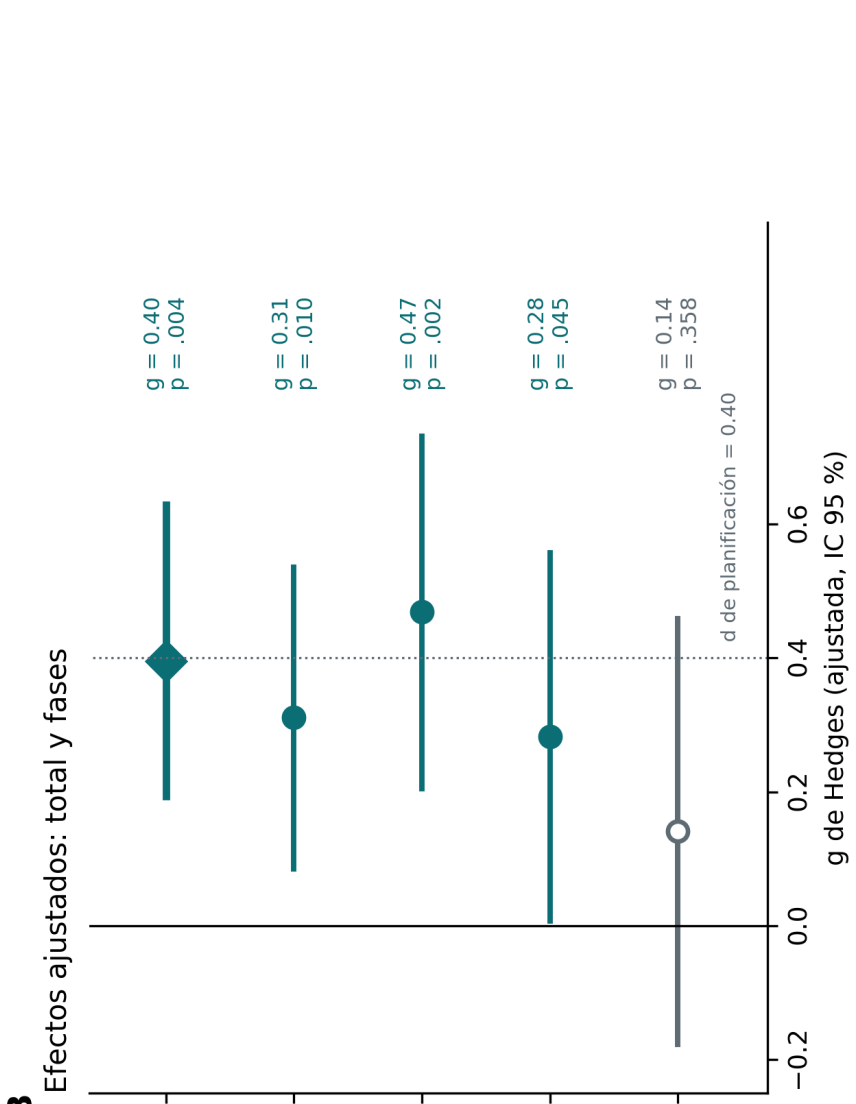
Nota: Cada modelo ajusta por el pretest de la misma fase, centrado en la media general. Las fases se contrastaron en secuencia fija (planificación, comprensión, ejecución, revisión), cada una con α = .05, tras rechazar H1 y hasta la primera no significativa. CCI = correlación intraclase.

Figura 11-2

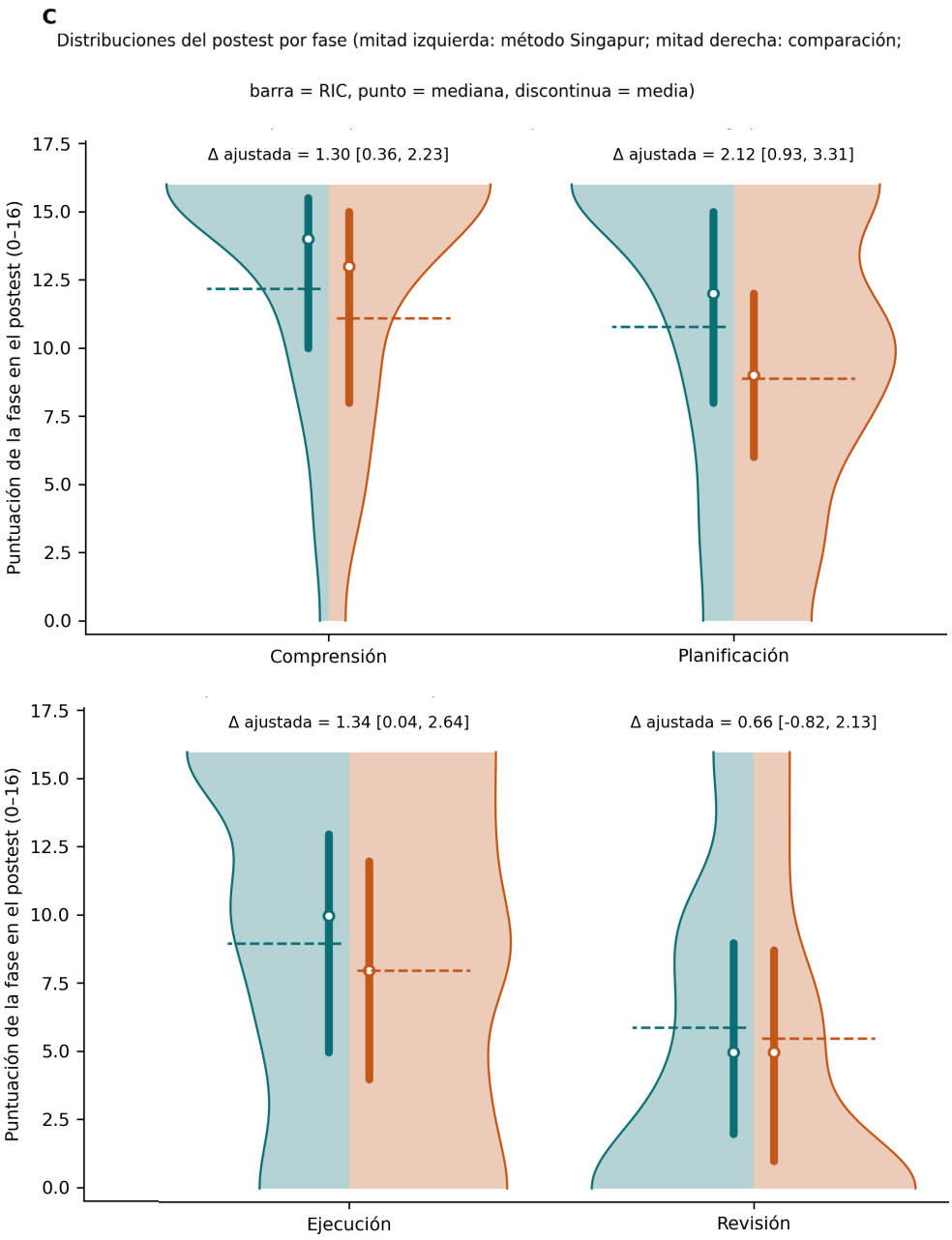
Cambio pretest–postest, efectos ajustados y distribuciones del postest por fase de resolución de problemas



Nota. Análisis de covarianza de dos niveles (estudiantes anidados en aulas, ajustados por el pretest; N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel A: medias del pretest (círculo) y del postest (punta de flecha) de cada fase por condición, con el IC 95 % del postest como banda; las etiquetas son ganancias medias (método Singapur: comprensión +2.98, planificación +3.85, ejecución +2.77, revisión +1.86; comparación: +1.19, +1.32, +0.97, +0.57). Fuente: Tablas 11-3, 11-5 y 11-6.



Nota. Análisis de covarianza de dos niveles (estudiantes anidados en aulas, ajustados por el pretest; N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel B: g de Hedges ajustada: total 0.40 (IC 95 % por bootstrap de conglomerados [0.19, 0.63], p = .004), comprensión 0.31 (p = .010), planificación 0.47 (p = .002), ejecución 0.28 (p = .045) y revisión 0.14 (p = .358); los intervalos de las fases son el IC 95 % del modelo para la diferencia ajustada expresado en unidades g, y los marcadores rellenos indican los efectos rechazados en la secuencia fija de pruebas que sigue al total (planificación, comprensión, ejecución, revisión; cada paso con $\alpha = .05$).
Fuente: Tablas 11-3, 11-5 y 11-6.



Nota. Análisis de covarianza de dos niveles (estudiantes anidados en aulas, ajustados por el pretest; N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel C: distribuciones del postest por fase (mitad izquierda método Singapur, mitad derecha comparación; barra = rango intercuartílico, punto = mediana, línea discontinua = media) con la diferencia ajustada y su IC 95 %: comprensión 1.30 [0.36, 2.23], planificación 2.12 [0.93, 3.31], ejecución 1.34 [0.04, 2.64] y revisión 0.66 [-0.82, 2.13]. Fuente: Tablas 11-3, 11-5 y 11-6.

11.4.7 Análisis de robustez

La Tabla 11-7 muestra que la estimación se mantuvo en las ocho especificaciones. Ignorar la conglomeración produjo la misma estimación puntual con un error estándar artificialmente pequeño. Los errores robustos a conglomerados, el bootstrap salvaje, el análisis por rangos de Quade y la estimación M de Huber respaldaron el efecto. Agregar sexo, edad e índice socioeconómico apenas modificó la estimación, y ninguna de estas covariables se asoció con el postest. El análisis sobre las 18 medias de aula y el modelo de puntuaciones de cambio arrojaron estimaciones algo mayores. Los estimados estandarizados oscilaron entre $g = 0.39$ y $g = 0.49$.

Tabla 11-7

Robustez del efecto de la condición sobre la puntuación total del postest

Especificación		Estimación	*EE*	Estadístico	*gl*	*p*	*g* de Hedges
ANCOVA de un nivel, *EE*	del modelo (ignora conglomerados)	6.04	1.18	$t = 5.13$	446	< .001	0.40
	robusto a conglomerados	6.04	1.69	$t = 3.58$	17	.002	0.40
Bootstrap-*t* de conglomerados salvaje (Rademacher, $B = 1999$)		6.04	1.69	$t = 3.58$	—	.003	0.40
ANCOVA	por rangos de Quade (robusto a conglomerados)	53.30 (rangos)	14.05	$t = 3.79$	17	.002	—
	con estimación M de Huber	6.35	1.25	$z = 5.07$	—	< .001	0.42
	de dos niveles con sexo, edad e índice socioeconómico	5.86	1.79	$t = 3.28$	15	.005	0.39
	a nivel de aula sobre 18 medias de aula	6.79	1.55	$t = 4.38$	15	< .001	0.45
Modelo de dos niveles de puntuaciones de cambio		7.39	1.55	$t = 4.76$	16	< .001	0.49

Nota: Estimaciones en puntos del postest, salvo el análisis de Quade (diferencia en residuos de rangos). En el modelo con covariables, sexo (b = 0.94, EE = 1.16), edad (b = −1.85, EE = 1.62) e índice socioeconómico (b = 0.54, EE = 0.79) no se asociaron con el postest. La g de Hedges usa la desviación típica combinada del postest.

11.4.8 Logro del dominio

Como indica la Tabla 11-8, las aulas del método Singapur, que comenzaron con menos estudiantes en el nivel de dominio, terminaron con más. La razón de momios ajustada por el pretest fue de 2.52, y las proporciones estandarizadas marginales, de 53.2 % y 35.1 %. A nivel de aula, la proporción de estudiantes con dominio aumentó en las nueve aulas del método Singapur (cambio medio de 28.0 puntos porcentuales) y en siete de las nueve de comparación (7.7 puntos). La Figura 11-3 muestra las distribuciones, la probabilidad predicha y las proporciones por aula.

Tabla 11-8

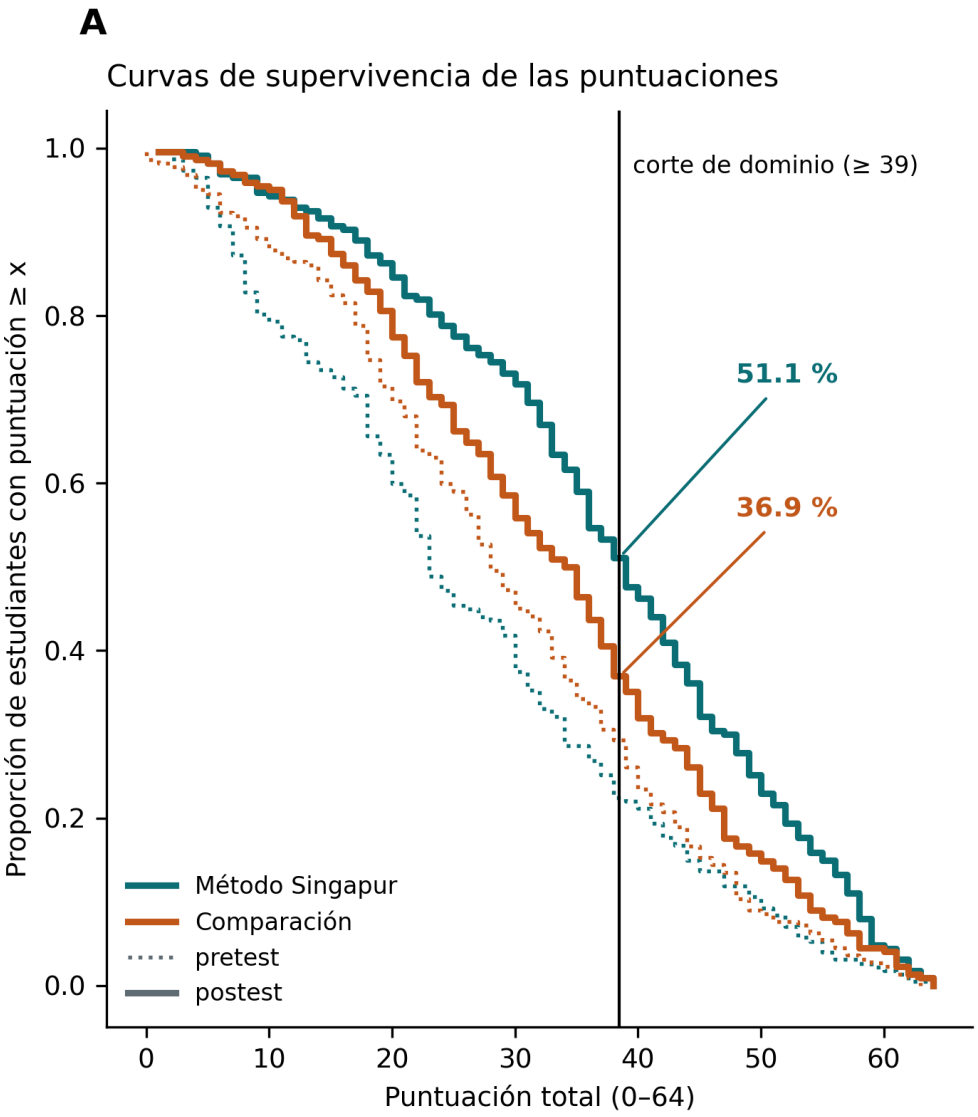
Proporción de estudiantes que alcanzan el dominio (total ≥ 39 de 64) según condición

Indicador		Método Singapur	Comparación	Contraste [IC 95 %]
Con dominio en	el pretest, %	22.5	29.3	—
	el postest, *n* (%)	116 (51.1)	82 (36.9)	—
Razón de momios ajustada (GEE, ajustada por el pretest)		—	—	2.52 [1.42, 4.49]
Proporción marginal estandarizada en el postest, %		53.2	35.1	—
Diferencia de riesgos, puntos porcentuales		—	—	18.2 [8.2, 28.0]
Número necesario a tratar		—	—	5.5

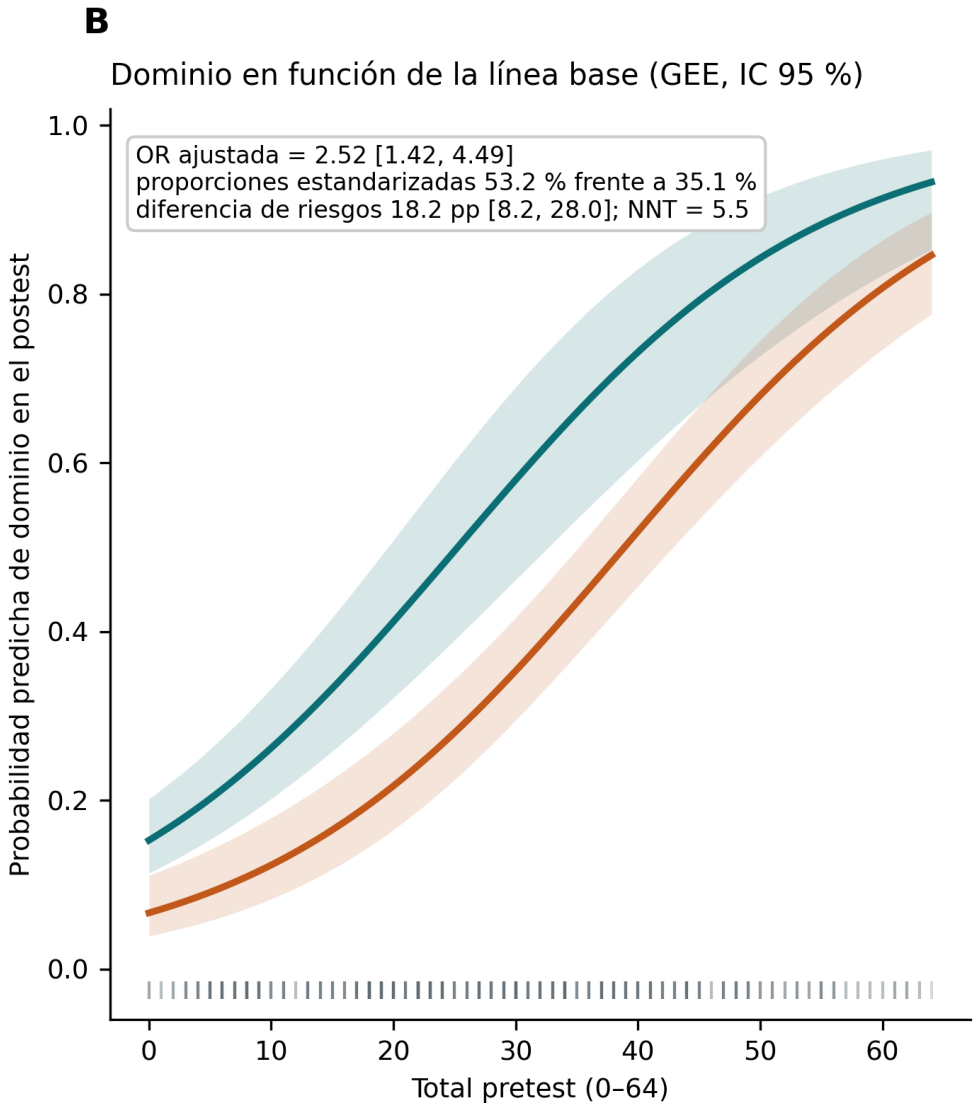
Nota: Ecuaciones de estimación generalizadas logísticas con correlación de trabajo intercambiable dentro de las aulas, ajustadas por el pretest total centrado en la media general. El IC de la diferencia de riesgos proviene de 1000 réplicas bootstrap de conglomerados.

Figura 11-3

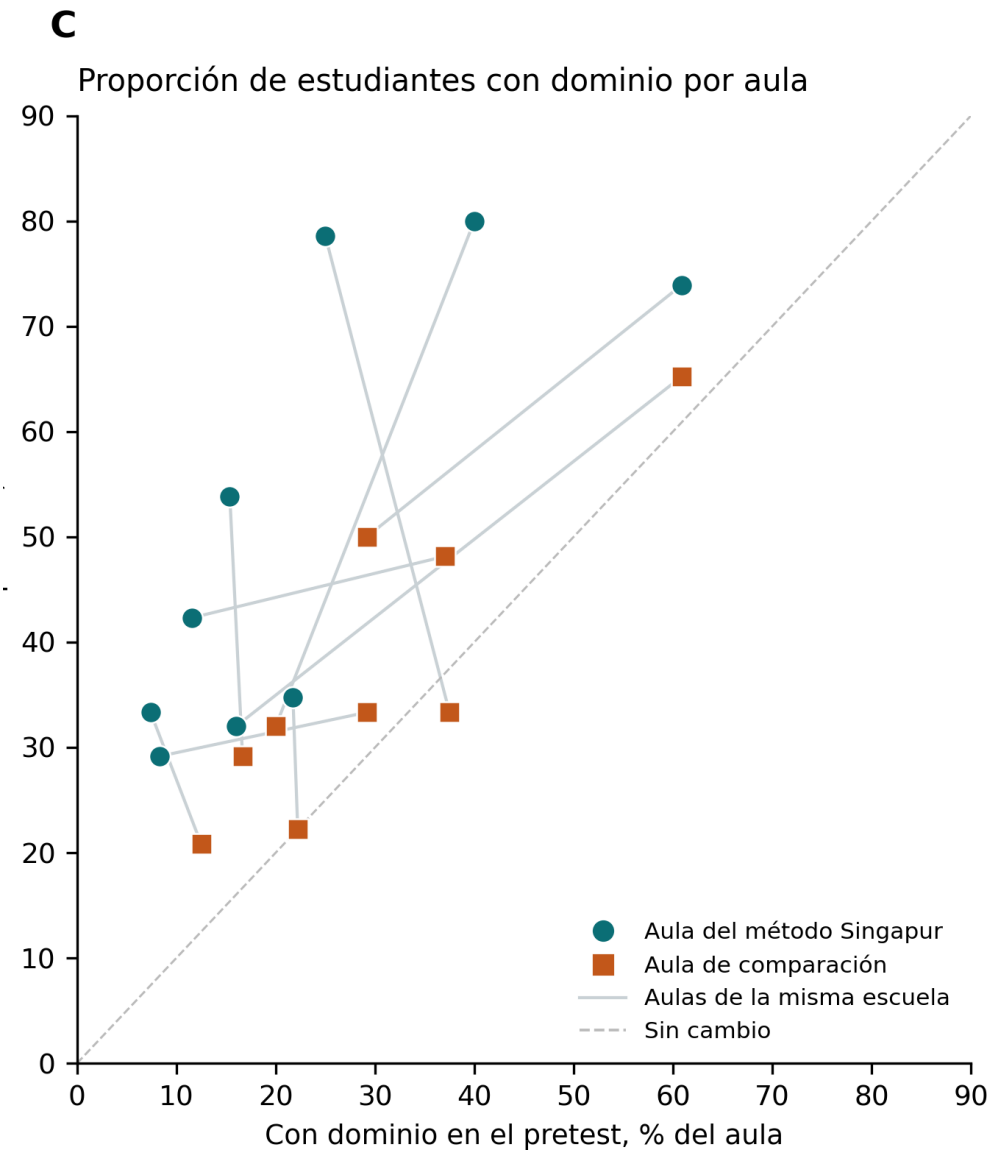
Distribuciones de las puntuaciones, probabilidad de dominio y proporción de estudiantes con dominio por aula



Nota. Dominio definido como un total en el postest ≥ 39; ecuaciones de estimación generalizadas logísticas con correlación intercambiable dentro de las aulas, ajustadas por el pretest (N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel A: proporción de estudiantes con una puntuación total igual o superior a cada valor en el pretest (punteada) y en el postest (continua); en el corte de dominio, el 51.1 % de los estudiantes del método Singapur y el 36.9 % de los de comparación alcanzaron el dominio en el postest. Fuente: Tabla 11-8 y puntuaciones por aula de la muestra analizada..



Nota. Dominio definido como un total en el postest ≥ 39; ecuaciones de estimación generalizadas logísticas con correlación intercambiable dentro de las aulas, ajustadas por el pretest (N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel B: probabilidad predicha de dominio en el postest en función de la puntuación del pretest según el modelo ajustado, con IC 95 %; OR ajustada = 2.52 (IC 95 % [1.42, 4.49]), proporciones estandarizadas 53.2 % y 35.1 %, diferencia de riesgos de 18.2 puntos porcentuales (IC 95 % por bootstrap de conglomerados [8.2, 28.0]) y número necesario a tratar (NNT) de 5.5, que expresan el efecto principal en la escala de dominio. Fuente: Tabla 11-8 y puntuaciones por aula de la muestra analizada.



Nota. Dominio definido como un total en el postest ≥ 39; ecuaciones de estimación generalizadas logísticas con correlación intercambiable dentro de las aulas, ajustadas por el pretest (N = 449 estudiantes en 18 aulas; método Singapur n = 227, comparación n = 222). Panel C: proporción de estudiantes con dominio en cada aula en el pretest (eje horizontal) y en el postest (eje vertical); las líneas grises unen las dos aulas de cada escuela y la diagonal discontinua marca la ausencia de cambio. La proporción aumentó en las nueve aulas del método Singapur (media 22.9 % → 50.9 %) y en siete de las nueve aulas de comparación (media 29.5 % → 37.1 %). Fuente: Tabla 11-8 y puntuaciones por aula de la muestra analizada..

11.4.9 Verificación de los parámetros de diseño

La Tabla 11-9 compara los valores de planificación con los estimados en los datos. La correlación intraclase observada (.143) fue cercana a la planificada, pero las correlaciones pretest-postest resultaron menores en ambos niveles, por lo que el factor de varianza fue 2.449 y no 1.922. El diseño realizado alcanzó una potencia del 80 % para efectos de $d \geq 0.44$, frente al $d = 0.40$ de planificación. La Figura 11-4 contrasta la sensibilidad planificada y la realizada y localiza el efecto en problemas y fases individuales.

Tabla 11-9

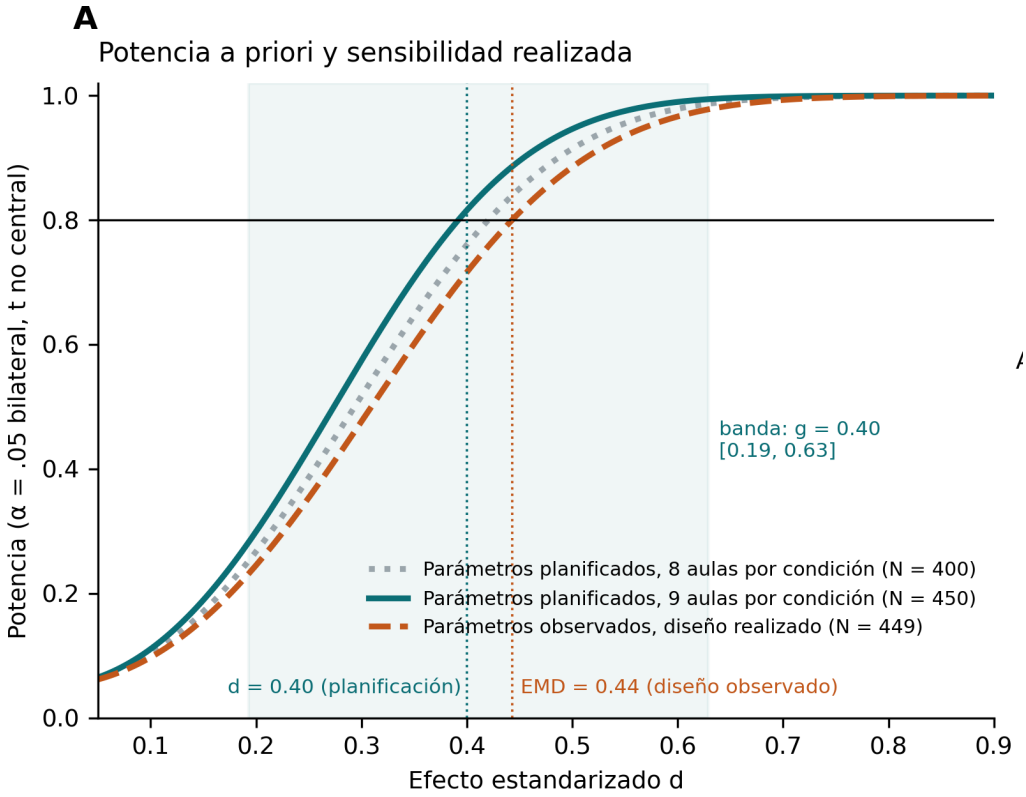
Parámetros de diseño planificados y observados

Parámetro			Planificado	Observado
Correlación	intraclase,	pretest total	.150	.143
		postest total	.150	.143
	pretest–postest a nivel de	estudiante (ρ_s)	.600	.506
		aula (ρ_c)	.800	.707
Tamaño medio de aula (m)			25.00	24.94
Correlación de las medias de aula, r (Teerenstra et al., 2012)			.763	.668
Efecto de diseño, $1 + (m - 1)\rho$			4.600	4.419
Factor de varianza, efecto de diseño × $(1 - r^2)$			1.922	2.449
Efecto mínimo detectable ($1 - \beta = .80$)			0.40	0.44

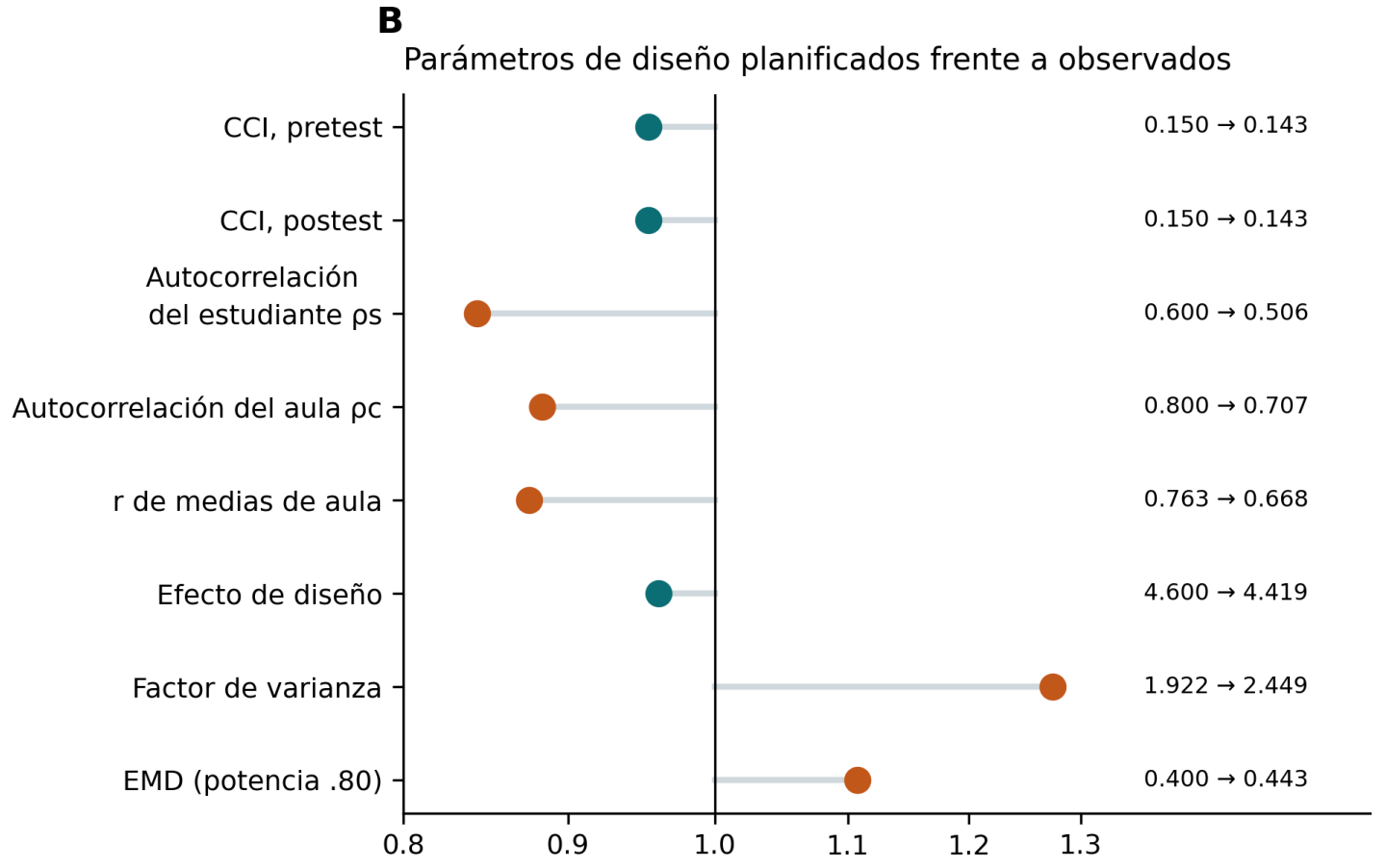
Nota: Correlaciones intraclase observadas a partir de modelos incondicionales de dos niveles; la correlación a nivel de estudiante usa puntuaciones centradas en las medias de aula. El efecto mínimo detectable planificado es el efecto para el que se calculó la muestra (aproximación normal, 7.54 aulas por condición); con las nueve aulas fijadas por la verificación con la t no central (gl = 15), los valores de planificación dan 0.39. El valor observado usa la distribución t no central con gl = 15.

Figura 11-4

Potencia planificada y realizada, parámetros de diseño y diferencias por problema y fase



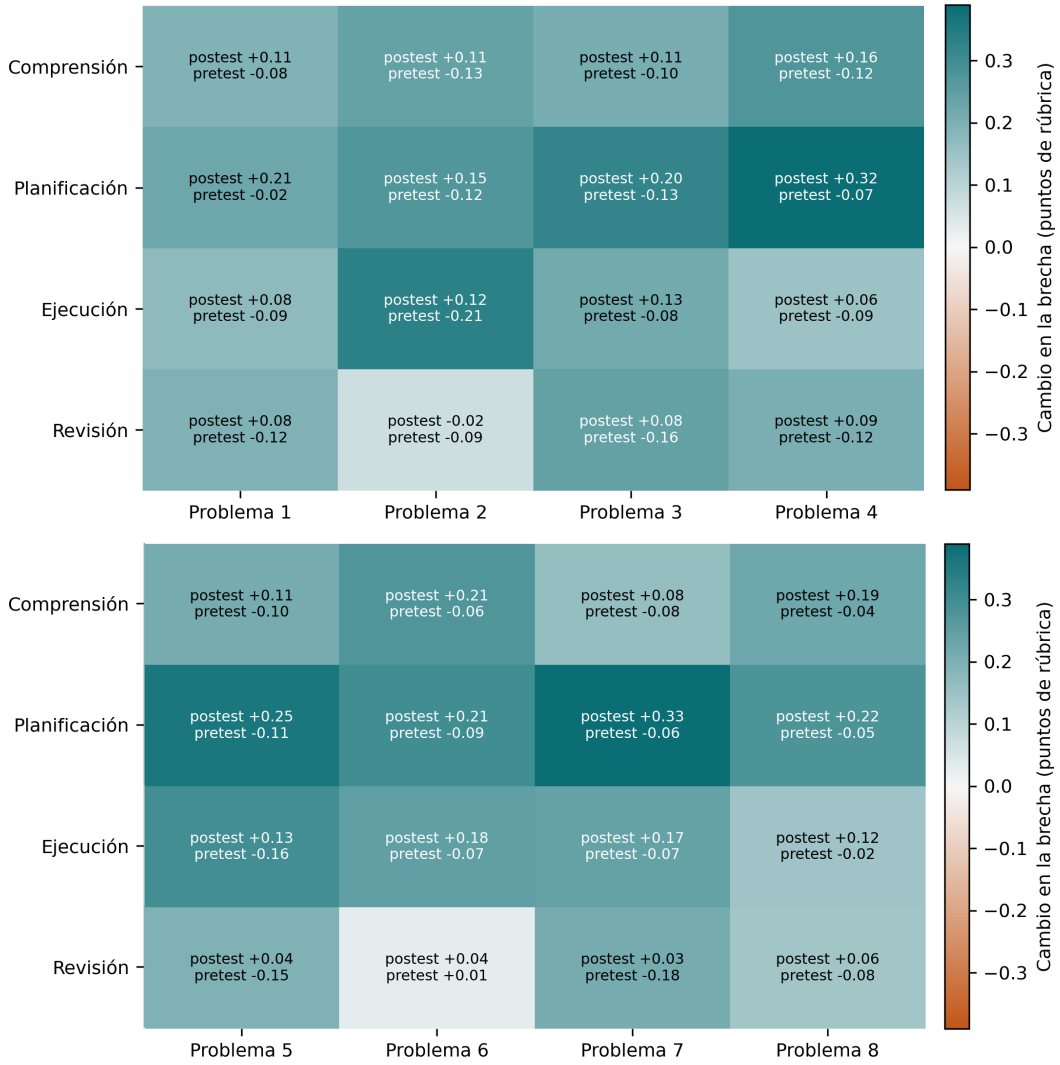
Nota. Potencia de la prueba del efecto de la condición calculada con la distribución t no central (α = .05 bilateral) para un diseño de dos niveles por conglomerados con ajuste por el pretest; diseño realizado N = 449 estudiantes en 18 aulas. Panel A: potencia con los parámetros de planificación con ocho aulas por condición (N = 400, gl = 13) y nueve aulas por condición (N = 450, gl = 15), y con los parámetros observados en el diseño realizado (N = 449, gl = 15); con nueve aulas por condición el efecto de planificación d = 0.40 alcanza una potencia de .816, y el diseño realizado tiene una potencia del 80 % para efectos d ≥ 0.44 (EMD = efecto mínimo detectable); la banda sombreada es la g observada = 0.40 (IC 95 % por bootstrap de conglomerados [0.19, 0.63]). Fuente: Tabla 11-9 y puntuaciones por ítem de la muestra analizada.



Nota. Potencia de la prueba del efecto de la condición calculada con la distribución t no central ($\alpha = .05$ bilateral) para un diseño de dos niveles por conglomerados con ajuste por el pretest; diseño realizado N = 449 estudiantes en 18 aulas. Panel B: razón observado/planificado de cada parámetro de diseño en escala logarítmica (planificado → observado: correlación intraclase (CCI) .150 → .143 en el pretest y en el postest, correlación pretest–postest a nivel de estudiante .600 → .506, a nivel de aula .800 → .707, correlación de las medias de aula .763 → .668, efecto de diseño 4.600 → 4.419, factor de varianza 1.922 → 2.449 y efecto mínimo detectable 0.400 → 0.443; el valor planificado es el efecto para el que se calculó la muestra, y con nueve aulas por condición y los parámetros de planificación el diseño detectaría d = 0.39). Fuente: Tabla 11-9 y puntuaciones por ítem de la muestra analizada.

C

Brecha por problema y fase (Singapur − comparación, puntos de rúbrica 0–2): diferencias en el postest y el pretest; color = postest − pretest



Nota. Potencia de la prueba del efecto de la condición calculada con la distribución t no central ($\alpha = .05$ bilateral) para un diseño de dos niveles por conglomerados con ajuste por el pretest; diseño realizado N = 449 estudiantes en 18 aulas. Panel C: para cada uno de los ocho problemas y cuatro fases, la diferencia método Singapur menos comparación en la puntuación media de la rúbrica (0–2) en el postest y en el pretest; el color muestra el cambio de esa brecha, que es mayor en la fase de planificación. Fuente: Tabla 9 y puntuaciones por ítem de la muestra analizada.

11.5 DISCUSIÓN

El hallazgo más desconcertante no es el efecto global, sino su ausencia en una sola fase: la revisión no se modificó ($g = 0.14$) pese a que las lecciones del grupo Singapur cerraban con una rutina explícita de cuatro pasos, uno de los cuales era precisamente verificar la respuesta. La expectativa razonable era que nombrar y practicar la verificación la mejorara; no ocurrió. Este resultado sugiere que enunciar el paso no basta y que la supervisión de la solución exige un diseño instruccional propio, en la línea de los enfoques de liberación gradual de la asistencia, que transfieren de manera progresiva el control heurístico al estudiante y refuerzan su repertorio heurístico y su confianza (Cerveza y Lapinid, 2024). Es consistente con ello que la enseñanza de heurísticas generales que explicitan el ciclo de solución se recomiende para los problemas verbales algebraicos (Witzel y Myers, 2023), lo que apunta a que la revisión se aprende como una práctica con andamiaje y retirada de apoyos, no como una línea más de una lista de pasos.

Con respecto a H1, el programa mejoró la resolución total de problemas en 0.40 desviaciones típicas tras ajustar por un inicio desfavorable y por la conglomeración, y elevó en 18 puntos porcentuales la proporción de estudiantes con dominio. Lo llamativo es que las aulas del método Singapur partieron por debajo y terminaron por encima; ese cruce difícilmente se explica por una regresión a la media, pues la estimación ajustada por el pretest, la de aula y la de cambio convergieron en la misma dirección. Conviene, sin embargo, leer la magnitud con cautela. Es claramente inferior a las medias de los metaanálisis de intervenciones en problemas verbales ($g = 0.71$ en estudiantes con dificultades matemáticas, Myers et al., 2023; $g = 0.95$ en primaria, Vessonen et al., 2025a), y esa distancia admite al menos tres lecturas no excluyentes. Los metaanálisis agregan sobre todo intervenciones administradas por investigadores, con mayor fidelidad y dosificación controlada, factores que moderan el efecto, mientras que aquí el programa lo impartieron docentes regulares a aulas completas; el efecto se estima frente a una instrucción habitual

que también dedicó el mismo tiempo a resolver problemas, y no frente a la ausencia de instrucción; y la unidad de asignación fue el aula, de modo que la precisión efectiva depende de dieciocho conglomerados y no de 449 individuos. Desde esta perspectiva, un efecto moderado medido en condiciones de campo no es un fracaso, sino el tipo de estimación que debería anclar las decisiones de escalamiento. Más aún, si se considera que una proporción importante de los efectos de los ensayos aleatorizados en educación con logros estandarizados es inferior a 0.05 desviaciones típicas (Kraft, 2023), el resultado se sitúa en el extremo favorable de lo que la literatura permite esperar de una intervención real. Ello no autoriza a ignorar que una prueba alineada con el currículo enseñado tiende a producir efectos mayores que una prueba estandarizada, y que la presente medición inmediata no informa sobre persistencia.

El perfil por fase constituye el aporte teórico central y respalda H2 de manera descriptiva. Los mayores efectos aparecieron en planificación y comprensión, las fases en que el estudiante debe representar cantidades y relaciones, que es lo que el modelo de barras externaliza (Ng, 2022); la ejecución mejoró menos, de forma compatible con un beneficio indirecto a través de mejores planes, y la revisión no cambió. Esta disociación favorece una explicación representacional del mecanismo frente a la de un simple aumento de práctica estructurada, que habría elevado todas las fases por igual. Debe matizarse que la diferencia entre comprensión y ejecución fue pequeña y que ninguna prueba contrastó las fases entre sí, de modo que la jerarquía es una ordenación de efectos y no una inferencia sobre diferencias. En un plano más especulativo, el liderazgo de la planificación podría relacionarse con las demandas ejecutivas de inhibición y flexibilidad cognitiva de esta fase (Schäfer et al., 2024), que un esquema visual aliviaría al reducir la carga de mantener las relaciones en la memoria; el estudio no midió funciones ejecutivas, por lo que esa conexión permanece como hipótesis.

Otro punto de contraste proviene de las intervenciones basadas en esquemas. Se ha mostrado que una intervención de esquemas con instrucción de comprensión del lenguaje cierra brechas de logro en

problemas verbales desde primer grado (Fuchs et al., 2021), y las transformaciones pictóricas que los estudiantes realizan sobre el modelo de barras (partición, completamiento y reposicionamiento) ofrecen una vía concreta por la que el esquema se convierte en plan (Richit et al., 2025). Nuestros datos no observaron esas transformaciones ni la dimensión lingüística, por lo que la coincidencia con estos mecanismos es de dirección y no de contenido. Tampoco se evaluó si el beneficio depende del rendimiento inicial, cuestión pertinente para el trabajo sobre manipulativos en distintos niveles de logro (Ahmad y Siller, 2024) y para los entornos con recursos limitados (Iyamuremye y Burns, 2025).

La evidencia regional sobre el método es más débil de lo que su difusión sugeriría. Los estudios disponibles proceden de escuelas únicas y grupos reducidos: un cuasiexperimento con grupos de control y experimental de 21 estudiantes de segundo grado en Perú (Gorbalan et al., 2025), un estudio preexperimental de problemas de igualación en cuarto grado en Perú (Jumpa et al., 2025) y una investigación con 37 estudiantes de quinto grado en Colombia (Rosero y Zambrano, 2025). Todos informan mejoras, pero con tamaños que impiden la inferencia a nivel de aula y sin tratar el anidamiento. La contribución de este trabajo es acotada y complementaria: aporta una estimación sensible a la conglomeración y resuelta por fase. Algo semejante cabe decir de las estrategias heurísticas evaluadas con un diseño preexperimental en 131 estudiantes de primaria en Perú (Vargas, 2025), cuya fortaleza descriptiva no sustituye un grupo de comparación.

La coincidencia con la literatura es, por último, más breve: la dirección del efecto concuerda con el papel atribuido al modelo de barras en la reducción de errores y con la revisión latinoamericana que lo vincula con la resolución de problemas (Baysal y Sevinç, 2022; Richit y Richit, 2022).

11.5.1 Implicaciones prácticas

Para los sistemas que adoptan currículos basados en representaciones, los resultados respaldan invertir en el núcleo representacional del método, la

secuencia concreto-pictórico-abstracta y el modelo de barras, e indican dónde este se queda corto. Dado que el modelo de barras resulta exigente también para los docentes (Lizano y Sevinç, 2025), la preparación debería incluir resolver y modelar problemas no triviales, no solo guiones de lección. Y puesto que la formulación de problemas recibe atención limitada en el propio currículo de Singapur (Toh y Chua, 2025), conviene no suponer que el método completo se transfiere sin adaptaciones. La ausencia de efecto en la revisión indica que la verificación requiere lecciones explícitas, por ejemplo, contrastar las respuestas con estimaciones o con la situación; enfoques metacognitivos de este tipo se han empleado con estudiantes de sexto grado con dislexia en problemas verbales algebraicos (Abdullah, 2023), lo que sugiere, con cautela, que la metacognición puede ser una vía para adaptar el método a perfiles de aprendizaje diversos. En cuanto a la implementación docente, la experiencia de programas de desarrollo profesional en matemáticas de 3.º a 5.º grado recuerda que los efectos sobre el logro dependen de cómo se sostiene el cambio en la práctica (Schoen et al., 2024), y en este estudio la fidelidad se observó pero no se relacionó con los resultados. Por último, la comparación de parámetros de diseño ofrece valores de planificación: futuras evaluaciones deberían asumir correlaciones pretest-postest menores que las empleadas aquí, y el diseño realizado solo detectaba con holgura efectos de 0.44.

11.5.2 Limitaciones y líneas futuras

El diseño fue cuasiexperimental: la condición siguió el horario escolar y las aulas de comparación partieron con puntuaciones más altas. El ANCOVA ajusta por el nivel inicial medido y el emparejamiento de aulas dentro de la escuela mantiene constantes las condiciones escolares, pero no pueden excluirse diferencias no medidas entre aulas, incluidas las características docentes. La inferencia descansa en 18 aulas y, aunque las correcciones para muestras pequeñas y el bootstrap salvaje coincidieron, el diseño realizado tenía potencia para 0.44 y no para 0.40. No se examinó la moderación por sexo ni por rendimiento inicial. La prueba se alineó con el currículo enseñado, el postest fue inmediato y no se modelaron

moderadores individuales asociados con el desempeño en problemas verbales (Vessonen et al., 2025b; Xu et al., 2024). Futuros estudios deberían combinar la asignación aleatoria de aulas, postests diferidos, medidas estandarizadas e instrucción explícita de la revisión.

11.6 CONCLUSIONES

Un programa de 12 semanas del método Singapur, impartido por docentes regulares, mejoró la resolución de problemas matemáticos de estudiantes de cuarto grado en 0.40 desviaciones típicas tras ajustar por el rendimiento inicial y por la conglomeración en aulas, y aumentó en 18 puntos porcentuales la proporción de estudiantes con dominio. El resultado se mantuvo en ocho especificaciones alternativas, de modo que no depende de una decisión analítica particular, y se obtuvo con una atrición baja y con instrumentos de alta fiabilidad.

El efecto se concentró en la planificación y la comprensión, las fases que el modelo de barras está diseñado para apoyar, fue menor en la ejecución y no alcanzó la revisión. Esta disociación sugiere que la instrucción basada en representaciones modifica la manera en que los estudiantes construyen y planifican las soluciones, pero que el control reflexivo de lo resuelto requiere una enseñanza propia, y no se activa por la mera inclusión de un paso de verificación en una rutina. Para la práctica, ello implica conservar el núcleo concreto-pictórico-abstracto y el modelo de barras, preparar a los docentes en la resolución y modelación de problemas no triviales y diseñar lecciones específicas de verificación.

Las conclusiones están acotadas por un diseño cuasiexperimental con 18 aulas, una medición inmediata y una prueba alineada con el currículo; en consecuencia, describen a las aulas participantes y deben leerse como una estimación de campo, no como un efecto garantizado en otros contextos. La comparación entre los parámetros de diseño planificados y los observados aporta, además, valores realistas para dimensionar futuras evaluaciones en aula. Una réplica con asignación aleatoria de aulas, seguimiento diferido y enseñanza explícita de la revisión permitiría establecer si el patrón por

fase aquí descrito es estable y si puede orientar la adaptación del método Singapur en sistemas educativos ajenos a Asia oriental.

11.7 REFERENCIAS BIBLIOGRÁFICAS

Abdullah, A. (2023). The experiences of Primary 6 students with dyslexia using the metacognitive-based approach of problem solving for algebraic word problems. *Asia Pacific Journal of Developmental Differences, 10*(1), 27-63. https://doi.org/10.3850/s2345734123000274

Ahmad, S., & Siller, H. (2024). Investigating the effect of manipulatives on mathematics achievement: The role of concrete and virtual manipulatives for diverse achievement level groups. *Journal on Mathematics Education, 15*(3), 979-1002. https://doi.org/10.22342/jme.v15i3.pp979-1002

Ayabe, H., Manalo, E., & Vries, E. (2022). Problem-appropriate diagram instruction for improving mathematical word problem solving. *Frontiers in Psychology, 13*, 992625. https://doi.org/10.3389/fpsyg.2022.992625

Baysal, E., & Sevinç, S. (2022). The role of the Singapore bar model in reducing students' errors on algebra word problems. *International Journal of Mathematical Education in Science and Technology, 53*(2), 289-310. https://doi.org/10.1080/0020739x.2021.1944683

Bell, S., Chalmers, R., & Flora, D. (2024). The Impact of Measurement Model Misspecification on Coefficient Omega Estimates of Composite Reliability. *Educational and Psychological Measurement, 84*(1), 5-39. https://doi.org/10.1177/00131644231155804

Cerveza, R., & Lapinid, M. (2024). The effects of gradual release of assistance instruction on students' heuristics, confidence, and attitude toward independent problem-solving. *Journal on Mathematics Education*, *15*(3), 771-792. https://doi.org/10.22342/jme.v15i3.pp771-792

Fuchs, L., Seethaler, P., Sterba, S., Craddock, C., Fuchs, D., Compton, D., Geary, D., & Changas, P. (2021). Closing the word-problem achievement gap in first grade: Schema-based word-problem intervention with embedded language comprehension instruction. *Journal of Educational Psychology*, *113*(1), 86-103. https://doi.org/10.1037/edu0000467

Gorbalan, I., Duran, K., & Mucha, L. (2025). Método Singapur para mejorar la resolución de problemas matemáticos en estudiantes de educación básica regular. *Episteme Koinonia*, *7*(1), 98-119. https://doi.org/10.35381/e.k.v7i1.4411

Hedges, L. (2025). Interpretation of the Standardized Mean Difference Effect Size When Distributions Are Not Normal or Homoscedastic. *Educational and Psychological Measurement*, *85*(2), 245-257. https://doi.org/10.1177/00131644241278928

Hedges, L., & Hedberg, E. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis*, *29*(1), 60-87. https://doi.org/10.3102/0162373707299706

Iyamuremye, E., & Burns, D. (2025). Concrete-Pictorial-Abstract instruction: enhancing students' learning motivation and achievement in mathematics. *Cogent Education*, *12*(1), 2558303. https://doi.org/10.1080/2331186x.2025.2558303

Jumpa, A., Huayra, M., & Rojas, A. (2025). El método Singapur en la Resolución de Problemas de Igualación en los Estudiantes de 4° Grado de la Institución Educativa N° 31030 Andaymarca. *Ciencia Latina Revista Científica Multidisciplinar, 9*(1), 1881-1899. https://doi.org/10.37811/cl_rcm.v9i1.15963

Kraft, M. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher, 49*(4), 241-253. https://doi.org/10.3102/0013189x20912798

Kraft, M. (2023). The Effect-Size Benchmark That Matters Most: Education Interventions Often Fail. *Educational Researcher, 52*(3), 183-187. https://doi.org/10.3102/0013189x231155154

Leong, Y., Toh, T., & Cheng, L. (2026). A revisitation of Concrete-Pictorial-Abstract. *Educational Studies in Mathematics, 122*(1), 1-22. https://doi.org/10.1007/s10649-025-10455-4

Lizano, C., & Sevinç, S. (2025). Examining preservice teachers' bar model representations for solving word problems involving fractions. *Journal of Mathematics Teacher Education.* https://doi.org/10.1007/s10857-025-09731-0

McNeish, D. (2026). A practical guide to selecting and blending approaches for clustered data: Clustered errors, multilevel models, and fixed-effect models. *Psychological Methods, 31*(2), 225-251. https://doi.org/10.1037/met0000620

Myers, J., Hughes, E., Witzel, B., Anderson, R., & Owens, J. (2023). A Meta-Analysis of Mathematical Interventions for Increasing the Word Problem Solving Performance of Upper Elementary

and Secondary Students with Mathematics Difficulties. *Journal of Research on Educational Effectiveness, 16*(1), 1-35. https://doi.org/10.1080/19345747.2022.2080131

Ng, S. (2022). The model method: Crown jewel in Singapore mathematics. *Asian Journal for Mathematics Education, 1*(2), 147-161. https://doi.org/10.1177/27527263221107526

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Pólya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press.

Richit, L., & Richit, A. (2022). O Modelo de Barras de Singapura na Resolução de Problemas Aritméticos e Algébricos. *Bolema: Boletim de Educação Matemática, 36*(73), 697-724. https://doi.org/10.1590/1980-4415v36n73a05

Richit, L., Stormowski, V., & Richit, A. (2025). Repertório estudantil de transformações pictóricas na resolução de problemas via Modelo de Barras de Singapura. *Praxis Educativa, 21*, 1-21. https://doi.org/10.5212/praxeduc.v.21.24648.002

Rosero, F., & Zambrano, R. (2025). Incidencia del método Singapur en resolución de problemas matemáticos en grado quinto. *Fedumar Pedagogía y Educación, 12*(1), 98-117. https://doi.org/10.31948/fpe.v12i1.4947

Saadati, F., Martínez, M., & Espinoza, C. (2025). Upper primary student attitudes toward mathematics problem solving; an exploratory study in Chile. *Research in Mathematics Education, 27*(1), 44-65. https://doi.org/10.1080/14794802.2023.2221659

Schoen, R., LaVenia, M., Tazaz, A., Gersten, R., & Smolkowski, K. (2024). Effects of a Mathematics Teacher Professional Development Program on Grades 3–5 Student Achievement. *The Elementary School Journal, 125*(2), 322-346. https://doi.org/10.1086/732783

Schäfer, J., Reuter, T., Leuchter, M., & Karbach, J. (2024). Executive functions and problem-solving—The contribution of inhibition, working memory, and cognitive flexibility to science problem-solving performance in elementary school students. *Journal of Experimental Child Psychology, 244,* 105962. https://doi.org/10.1016/j.jecp.2024.105962

Teerenstra, S., Eldridge, S., Graff, M., De Hoop, E., & Borm, G. (2012). A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine, 31*(20), 2169-2178. https://doi.org/10.1002/sim.5352

Toh, T., & Chua, P. (2025). Problem posing in the Singapore mathematics classroom: A review of the Singapore mathematics curriculum. *Asian Journal for Mathematics Education, 4*(2), 256-279. https://doi.org/10.1177/27527263251339205

Vargas, M. (2025). Estrategias heurísticas en la resolución de problemas matemáticos en una institución educativa primaria en Perú. *Revista Innova Educación, 7*(1), 36-44. https://doi.org/10.35622/j.rie.2025.01.003

Vessonen, T., Dahlberg, M., Hellstrand, H., Widlund, A., Korhonen, J., Aunio, P., & Laine, A. (2024). Task Characteristics Associated with Mathematical Word Problem-Solving Performance Among Elementary School-Aged Children: A Systematic Review and Meta-Analysis. *Educational Psychology Review*, *36*(4), 117. https://doi.org/10.1007/s10648-024-09954-2

Vessonen, T., Dahlberg, M., Hellstrand, H., Widlund, A., Söderberg, P., Korhonen, J., Aunio, P., & Laine, A. (2025b). Individual characteristics associated with elementary school children's mathematical word problem-solving skills: A systematic review and meta-analysis. *Review of Education*, *13*(1), e70045. https://doi.org/10.1002/rev3.70045

Vessonen, T., Hellstrand, H., Kurkela, M., Aunio, P., & Laine, A. (2025a). The effectiveness of mathematical word problem-solving interventions among elementary schoolers: A systematic review and meta-analysis. *International Journal of Educational Research*, *132*, 102642. https://doi.org/10.1016/j.ijer.2025.102642

Witzel, B., & Myers, J. (2023). Solving Algebraic Word Problems Using General Heuristics Instruction. *TEACHING Exceptional Children*, *56*(1), 52-60. https://doi.org/10.1177/00400599231157029

Xu, C., Di Lonardo, S., Li, H., Liu, C., & Si, J. (2024). From whole numbers to fractions to word problems: Hierarchical relations in mathematics knowledge for Chinese Grade 6 students. *Journal of Experimental Child Psychology*, *242*, 105884. https://doi.org/10.1016/j.jecp.2024.105884

Yamamoto, Y., & Miyazaki, Y. (2025). Small Sample Methods in Multilevel Analysis. *The Journal of Experimental Education, 93*(1), 186-217. https://doi.org/10.1080/00220973.2024.2329084

CAPÍTULO XII

LA EXCURSIÓN GEOGRÁFICA COMO ESTRATEGIA DIDÁCTICA VIVENCIAL Y EL RENDIMIENTO ACADÉMICO EN GEOGRAFÍA: EL PAPEL MEDIADOR DE LA COMPRENSIÓN ESPACIAL EN ESTUDIANTES DE SECUNDARIA DEL PERÚ, 2025

Mauricio Acevedo Carrillo

Universidad Privada del Norte

Resumen

La excursión geográfica es una de las estrategias más antiguas de la enseñanza de la geografía, pero la evidencia sobre cómo se traduce en aprendizaje escolar sigue siendo fragmentaria y rara vez contrasta el mecanismo cognitivo implicado. Este estudio examina si la experiencia de excursión geográfica vivencial se asocia con el rendimiento académico en geografía de estudiantes de secundaria y si esa asociación se transmite a través de la comprensión espacial. Se aplicó un diseño explicativo transversal con 492 estudiantes de tercero a quinto de secundaria de instituciones públicas del Perú, cuyo tamaño se fijó a priori para una potencia de .90 en la prueba de ajuste cercano. El modelo se estimó con ecuaciones estructurales por máxima verosimilitud y el efecto indirecto se contrastó con intervalos bootstrap percentil de 5000 réplicas. El modelo de medida mostró cargas entre .720 y .833 y fiabilidades compuestas entre .775 y .854, y el modelo estructural se ajustó a los datos, $\chi^2(24) = 16.831$, $p = .856$, CFI = 1.000, RMSEA = .000, SRMR = .016. La experiencia de excursión se asoció con la comprensión espacial ($\beta = .751$) y con el rendimiento de forma directa ($\beta = .376$); la comprensión espacial se asoció con el rendimiento ($\beta = .316$), y el efecto indirecto fue de .237, IC 95 % [.088, .379], el 38.7 % del efecto total (.613). Los resultados son compatibles con una mediación parcial en la que la excursión se asocia con el rendimiento por la vía del razonamiento espacial y por vías no espaciales, y sitúan la planificación de la salida —objetivos, registro y reflexión posterior— como el núcleo didáctico que conviene preservar cuando la excursión se reduce o se sustituye.

Palabras clave: excursión geográfica; aprendizaje vivencial; comprensión espacial; pensamiento espacial; rendimiento académico; educación secundaria; ecuaciones estructurales; Perú.

Abstract

Geographic fieldwork is among the oldest strategies in geography teaching, yet evidence on how it translates into school learning remains fragmentary and

seldom tests the cognitive mechanism involved. This study examines whether experiential geographic field trip experience is associated with secondary students' academic achievement in geography, and whether that association is transmitted through spatial understanding. An explanatory cross-sectional design was applied to 492 third- to fifth-year students from public schools in Peru, with a sample size fixed a priori to reach a power of .90 for the close-fit test. The model was estimated with maximum-likelihood structural equation modeling, and the indirect effect was tested with percentile bootstrap intervals from 5000 resamples. The measurement model showed loadings between .720 and .833 and composite reliabilities between .775 and .854, and the structural model fitted the data, $\chi^2(24) = 16.831$, $p = .856$, CFI = 1.000, RMSEA = .000, SRMR = .016. Field trip experience was associated with spatial understanding ($\beta = .751$) and directly with achievement ($\beta = .376$); spatial understanding was associated with achievement ($\beta = .316$), and the indirect effect was .237, 95% CI [.088, .379], 38.7% of the total effect (.613). The results are compatible with partial mediation, in which the field trip is associated with achievement through spatial reasoning and through non-spatial routes, and they identify trip planning —objectives, recording, and post-trip reflection— as the didactic core worth preserving when the excursion is shortened or replaced.

Keywords: geographic field trip; experiential learning; spatial understanding; spatial thinking; academic achievement; secondary education; structural equation modeling; Peru.

12.1 INTRODUCCIÓN

La salida de campo ocupa un lugar singular en la enseñanza de la geografía: es el momento en que el territorio deja de ser una representación —un mapa, una imagen, un dato— y se convierte en objeto de observación directa. Las revisiones recientes coinciden en que la enseñanza fuera del aula mejora el aprendizaje escolar (Pulido et al., 2025), en que la salida rinde cuando se prepara con una etapa previa de planificación (Menezes, 2025; Evangelista et al., 2026) y en que puede servir para trabajar contenidos tan diversos como la geografía política (Megoran y Benwell, 2026) o

integrarse en prácticas más amplias de educación al aire libre (Aylward et al., 2026; Remmen y Iversen, 2023). Al mismo tiempo, la excursión está bajo presión: su coste, sus barreras de equidad y su huella ambiental la han vuelto discutible en términos de sostenibilidad (Woodley et al., 2024), la inteligencia artificial generativa empieza a transformar el trabajo de campo (Lee et al., 2025) y las salidas virtuales inmersivas mejoran el rendimiento en ciencias sociales (Alazmi y Alemtairy, 2024).

La literatura reciente ha avanzado en dos direcciones que rara vez se cruzan. La primera documenta los efectos de la salida de campo y de sus variantes: evaluaciones de salidas en ciencias (Dobečková y Maršálek, 2026), modelos de trabajo de campo combinados con aprendizaje en línea (Oktavianto et al., 2024b), propuestas que articulan el campo con tecnologías de la información (De Sousa y Barbosa, 2025) y asignaciones de roles que sostienen la implicación del estudiante durante la salida (Minor y McCourt, 2021). La segunda estudia el pensamiento espacial como competencia educable: el entrenamiento mejora las habilidades espaciales, con un efecto medio de g = 0.47 en 217 estudios (Uttal et al., 2013), y estas se asocian con el rendimiento matemático, r = .36 en 45 artículos (Atit et al., 2022). En geografía, el pensamiento espacial se ha descompuesto en componentes evaluables (Lee y Bednarz, 2012; Tserendolgor y Choijilsuren, 2026), sus determinantes se han modelado con ecuaciones estructurales (Xie et al., 2021), sus procesos cognitivos difieren entre estudiantes (Xiang y Xi, 2025; Liang et al., 2026) y se ha mostrado sensible a intervenciones con sistemas de información geográfica (Cao et al., 2024; Rodriguez y Aguilar, 2026), con cómics (Von Reumont y Budke, 2021) y con aprendizaje basado en proyectos (Wei y Zhao, 2026).

Lo que falta es el puente entre ambas líneas. Si la excursión mejora el aprendizaje geográfico, la pregunta científica no es solo si lo hace, sino por qué vía. La hipótesis más coherente con la teoría del aprendizaje experiencial es que la observación directa del territorio, el registro guiado y la reflexión posterior consolidan relaciones de localización, escala y distribución que el estudiante luego moviliza al explicar problemas territoriales; es decir, que la comprensión espacial transmitiría una

parte del efecto de la excursión sobre el rendimiento. Esa hipótesis de mecanismo no se ha contrastado con un modelo estructural que separe el efecto directo del indirecto en estudiantes de secundaria. En los sistemas escolares latinoamericanos, la investigación sobre el trabajo de campo se ha centrado en experiencias formativas, en la salida pedagógica y en las concepciones de quienes la diseñan (De Sousa y Sepúlveda, 2024; Mantilla et al., 2025; Retana y Cascante, 2026), y la investigación sobre el pensamiento geográfico y espacial se ha expresado sobre todo en ensayos, revisiones y propuestas didácticas (Moreno et al., 2023; Andreis y Claudino, 2024; Araya y Álvarez, 2025, 2026; Guallart, 2025; Haquet y Aguilera, 2025; Romero, 2025; Rosero y Gómez, 2025; Díaz y Galán, 2026; Quintero, 2026), a menudo vinculadas a una concepción de la geografía como conocimiento poderoso y como ciudadanía espacial (Cascante, 2024; Crane y Mewborne, 2026). Los diagnósticos disponibles, por su parte, muestran porcentajes de acierto muy reducidos en las pruebas de localización (Gómez et al., 2025).

El Perú es un sitio de prueba exigente para esta pregunta. Su currículo nacional organiza la geografía escolar en torno a la competencia «Gestiona responsablemente el espacio y el ambiente» (Ministerio de Educación del Perú, 2016), que exige comprender las relaciones entre los elementos naturales y sociales del espacio, manejar fuentes de información geográfica y proponer acciones ante problemáticas territoriales; su diversidad de paisajes ofrece a cada escuela un territorio inmediato observable; y sus resultados de aprendizaje siguen por debajo del promedio de la OCDE (OECD, 2023), con rendimientos en ciencias sociales que, en la secundaria rural, no se explican por la preparación docente, la enseñanza ni la gestión institucional (Chocce, 2026). La contribución del estudio, sin embargo, no es regional. El modelo que se contrasta —experiencia de campo, comprensión espacial, rendimiento— es un modelo general del aprendizaje geográfico experiencial, y su prueba en un sistema escolar con recursos limitados permite discutir qué componente de la excursión debe preservarse cuando la salida se acorta, se reduce o se reemplaza.

El objetivo es estimar la asociación de la experiencia de excursión geográfica vivencial con el rendimiento académico en geografía de estudiantes de secundaria del Perú y contrastar si esa asociación se transmite a través de la comprensión espacial. La pregunta de investigación es: ¿en qué medida la comprensión espacial media la relación entre la experiencia de excursión geográfica vivencial y el rendimiento académico en geografía? Se formulan cuatro hipótesis, desarrolladas en la sección siguiente: un efecto directo de la excursión sobre el rendimiento (H1), un efecto de la excursión sobre la comprensión espacial (H2), un efecto de la comprensión espacial sobre el rendimiento (H3) y un efecto indirecto de la excursión sobre el rendimiento a través de la comprensión espacial (H4).

12.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

12.2.1 La excursión geográfica como experiencia de aprendizaje

La excursión geográfica se entiende como una actividad de aprendizaje fuera del aula, planificada dentro del área curricular, en la que el estudiante observa directamente fenómenos del territorio, registra información con instrumentos de campo y reflexiona después sobre lo observado. Desde la teoría del aprendizaje experiencial, el valor de la salida no reside en el desplazamiento, sino en la secuencia de experiencia concreta, observación reflexiva y conceptualización. La investigación clásica mostró que el aprendizaje depende de la calidad de la salida y de la familiaridad previa con el entorno (Orion y Hofstein, 1994), principio que la literatura reciente concreta en una etapa de pre-campo (Menezes, 2025; Evangelista et al., 2026). Por eso este estudio no mide la mera asistencia a una salida, sino la calidad de la experiencia vivencial que el estudiante reporta: observación directa, objetivos y registro, y reflexión posterior.

12.2.2 La comprensión espacial como mecanismo

La comprensión espacial es la capacidad de interpretar relaciones de localización y orientación, de comprender los efectos de la escala en la representación de los fenómenos y de reconocer patrones de distribución

en mapas, imágenes y paisajes. Corresponde al núcleo del pensamiento espacial en geografía (Lee y Bednarz, 2012) y es una capacidad educable, que mejora con el entrenamiento y la experiencia (Uttal et al., 2013). La salida de campo es una situación privilegiada para ese aprendizaje porque obliga a orientarse, a estimar distancias y a comparar el territorio con su representación cartográfica, y puede modificar las concepciones espaciales de los futuros docentes (Ulbricht y Lindau, 2025).

12.2.3 Del pensamiento espacial al rendimiento académico

El rendimiento académico en geografía se operacionaliza como el logro de la competencia curricular de gestión del espacio y el ambiente: explicar las relaciones entre los elementos que configuran el espacio, usar mapas, gráficos y datos para analizar problemáticas territoriales y proponer acciones fundamentadas (Ministerio de Educación del Perú, 2016). Esa competencia descansa sobre operaciones espaciales —localizar, comparar escalas, reconocer distribuciones—, por lo que cabe esperar que la comprensión espacial se asocie con ella, del mismo modo en que las habilidades espaciales se asocian con el rendimiento matemático (Atit et al., 2022).

12.2.4 Hipótesis

El modelo propuesto sostiene que la excursión aporta al rendimiento por dos vías: una indirecta, a través de la comprensión espacial, y una directa que recoge otros mecanismos de la experiencia de campo, como la motivación, la conexión con el entorno o la consolidación de contenidos no espaciales. Se formulan cuatro hipótesis:

- **H1.** La experiencia de excursión geográfica vivencial se asocia de forma directa y positiva con el rendimiento académico en geografía (ruta c′).
- **H2.** La experiencia de excursión geográfica vivencial se asocia de forma positiva con la comprensión espacial (ruta a).

- **H3.** La comprensión espacial se asocia de forma positiva con el rendimiento académico en geografía, controlando la experiencia de excursión (ruta b).
- **H4.** La experiencia de excursión geográfica vivencial tiene un efecto indirecto positivo sobre el rendimiento académico en geografía a través de la comprensión espacial (efecto a·b).

H1, H2 y H3 se deciden por el coeficiente estandarizado de su ruta y su intervalo de confianza; H4, por el intervalo bootstrap percentil del efecto indirecto.

12.3 MATERIALES Y MÉTODOS

12.3.1 Diseño y participantes

Se aplicó un diseño cuantitativo, no experimental, explicativo y transversal, con un modelo de ecuaciones estructurales basado en covarianzas que incluyó tres constructos latentes —experiencia de excursión geográfica vivencial (exógena), comprensión espacial (mediador) y rendimiento académico en geografía (endógena final)— y tres rutas (a, b y c′). Los datos se recogieron en el año escolar 2025. Participaron 492 estudiantes de tercero a quinto de secundaria de instituciones públicas del Perú que habían realizado al menos una excursión geográfica en el área de Ciencias Sociales ese año. La muestra fue no probabilística por conveniencia, de modo que los resultados describen a los estudiantes analizados y no a la población escolar peruana.

12.3.2 Tamaño de la muestra

El tamaño se fijó antes de la recogida con tres criterios, uno por análisis, y se adoptó el más exigente (Tabla 1; Figura 4A).

1) El primero fue la potencia de la prueba de ajuste cercano (MacCallum et al., 1996): con 9 indicadores y 21 parámetros libres, el modelo tiene

24 grados de libertad; se contrastó H_0: ε = .05 frente a H_a: ε = .08 con α = .05.

2) El segundo fue la prueba F parcial de la ruta b, anclada en r = .36 (Atit et al., 2022), con f^2 = .149.

3) El tercero fue la potencia del efecto indirecto, estimada por simulación de Monte Carlo con intervalo de confianza del 95 % (Schoemann et al., 2017), con la ruta a anclada en g = 0.47 (Uttal et al., 2013), equivalente a r = .229, y la ruta b en r = .36, con 1000 réplicas por tamaño.

El criterio más exigente fue el ajuste cercano con potencia de .90, por lo que N = 492, tamaño que cubre los tres análisis.

Tabla 12-1

Cálculo a priori del tamaño de la muestra

Criterio	Análisis	Parámetros	Potencia objetivo	N mínimo	Potencia con N = 492
Ajuste cercano (RMSEA)	Modelo de ecuaciones estructurales	gl = 24; ε_0 = .05; ε_a = .08; α = .05	.80	375	.901
		gl = 24; ε_0 = .05; ε_a = .08; α = .05	.90	492	.901
F parcial de la ruta b	Ecuación del rendimiento (2 predictores)	f^2 = .149 (r = .36); α = .05	.80	56	> .999
		f^2 = .149 (r = .36); α = .05	.90	74	> .999
Efecto indirecto a·b	Mediación (Monte Carlo, IC 95 %)	a = .229; b = .36; 1000 réplicas	.80	150	.999
		a = .229; b = .36; 1000 réplicas	.90	210	.999

Nota: Ajuste cercano: prueba de H_0: ε = .05 frente a H_a: ε = .08 con λ = (N − 1)·gl·ε^2 (MacCallum et al., 1996). F parcial: λ = f^2·(u + v + 1), u = 1, v = N − 3. Efecto indirecto: simulación de Monte Carlo con intervalo de confianza de Monte Carlo (Schoemann et al., 2017). Anclas del tamaño del efecto: g = 0.47 en 217 estudios (Uttal et al., 2013) y r = .36 en 45 artículos (Atit et al., 2022). El N adoptado (492) es el mínimo del criterio más exigente y cubre los tres análisis.

12.3.3 Instrumentos y procedimiento

Los tres constructos se midieron con escalas de elaboración propia de tres ítems cada una, con respuesta tipo Likert de cinco puntos. Los ítems de la experiencia de excursión se redactaron a partir de los factores que condicionan el aprendizaje en una salida de campo (Orion y Hofstein, 1994); los de la comprensión espacial, a partir de los componentes del pensamiento espacial (Lee y Bednarz, 2012); y los del rendimiento, a partir de las capacidades de la competencia «Gestiona responsablemente el espacio y el ambiente» (Ministerio de Educación del Perú, 2016). La Tabla 12-2 presenta los ítems y las propiedades del modelo de medida. El cuestionario se aplicó de forma presencial, en el aula y en horario escolar, tras la última excursión del año de cada grupo; la participación fue voluntaria y anónima, y los bloques de ítems de cada constructo se presentaron por separado.

Tabla 12-2

Ítems y propiedades del modelo de medida

Constructo	Ítem	λ	R^2	α	ω	AVE
Excursión geográfica vivencial	Durante la excursión observé directamente los fenómenos del territorio estudiados en clase	.720	.518	.775	.775	.535
	Las actividades de campo tuvieron objetivos de aprendizaje y registro claros (fichas, croquis, mapas)	.724	.524	.775	.775	.535
	La reflexión posterior a la salida me permitió relacionar el territorio con los conceptos de geografía	.749	.561	.775	.775	.535
Comprensión espacial	Interpreto relaciones de localización y orientación entre elementos del territorio	.773	.598	.854	.854	.661
	Comprendo cómo cambia la representación de un fenómeno según la escala del mapa	.832	.692	.854	.854	.661
	Identifico patrones de distribución espacial en mapas, imágenes y paisajes	.833	.694	.854	.854	.661

Tabla 1-3

Propiedades psicométricas del modelo de medición (análisis factorial confirmatorio)

Constructo	Ítem	λ	R^2	α	ω	AVE
Rendimiento académico en geografía	Explico las relaciones entre los elementos naturales y sociales que configuran el espacio geográfico	.769	.591	.835	.836	.630
	Utilizo mapas, gráficos y datos para analizar problemáticas territoriales y ambientales	.832	.692	.835	.836	.630
	Propongo acciones fundamentadas ante problemáticas del territorio de mi localidad	.778	.605	.835	.836	.630

Nota: N = 492. λ = carga factorial estandarizada del análisis factorial confirmatorio; R^2 = varianza del ítem explicada por su constructo; α = alfa de Cronbach; ω = fiabilidad compuesta omega; AVE = varianza media extraída. Criterios de referencia: λ ≥ .50; α y ω ≥ .70; AVE ≥ .50..

12.3.4 Análisis de datos

1) Primero se evaluó el modelo de medida: cargas estandarizadas, fiabilidad (α y ω), validez convergente (AVE) y validez discriminante, con la razón heterorrasgo-monorrasgo (HTMT) y umbral de .85 (Henseler et al., 2015) y con el criterio según el cual la raíz cuadrada de la AVE debe superar las correlaciones con los demás constructos (Fornell y Larcker, 1981).

2) Segundo, el modelo estructural se estimó por máxima verosimilitud y su ajuste se evaluó con χ^2, CFI, TLI, RMSEA con IC del 90 %, prueba de ajuste cercano y SRMR, con valores de referencia próximos a .95, .06 y .08 para CFI/TLI, RMSEA y SRMR (Hu y Bentler, 1999); se comparó con un modelo alternativo de un solo factor mediante la diferencia de χ^2. Las rutas, el efecto indirecto (a·b) y el efecto total ($c = c' + a \cdot b$) se estimaron con intervalos bootstrap percentil del 95 % con 5000 réplicas; un efecto se consideró distinto de cero si su intervalo excluía el cero. La varianza explicada y el tamaño del efecto por ruta (f^2) se derivaron de los coeficientes estandarizados. Como sensibilidad, las rutas se re-estimaron por mínimos cuadrados sobre puntuaciones compuestas, y

la estabilidad del efecto total se evaluó con jackknife, submuestreos del 80 % y 60 %, bootstrap, winsorización al 5 % del rendimiento y una prueba placebo con la excursión permutada.

12.4 RESULTADOS

12.4.1 Estadísticos descriptivos y correlaciones

La Tabla 12-3 presenta los estadísticos descriptivos y las correlaciones de los tres constructos, y la Figura 12-1 muestra su distribución conjunta, el perfil por deciles de la excursión y la distribución de cada constructo. Las medias se situaron en el punto central de la escala, y la asimetría y la curtosis fueron pequeñas (|asimetría| ≤ 0.214; |curtosis| ≤ 0.100). Las tres correlaciones fueron positivas y significativas; la más alta correspondió a la excursión con la comprensión espacial.

Tabla 12-3

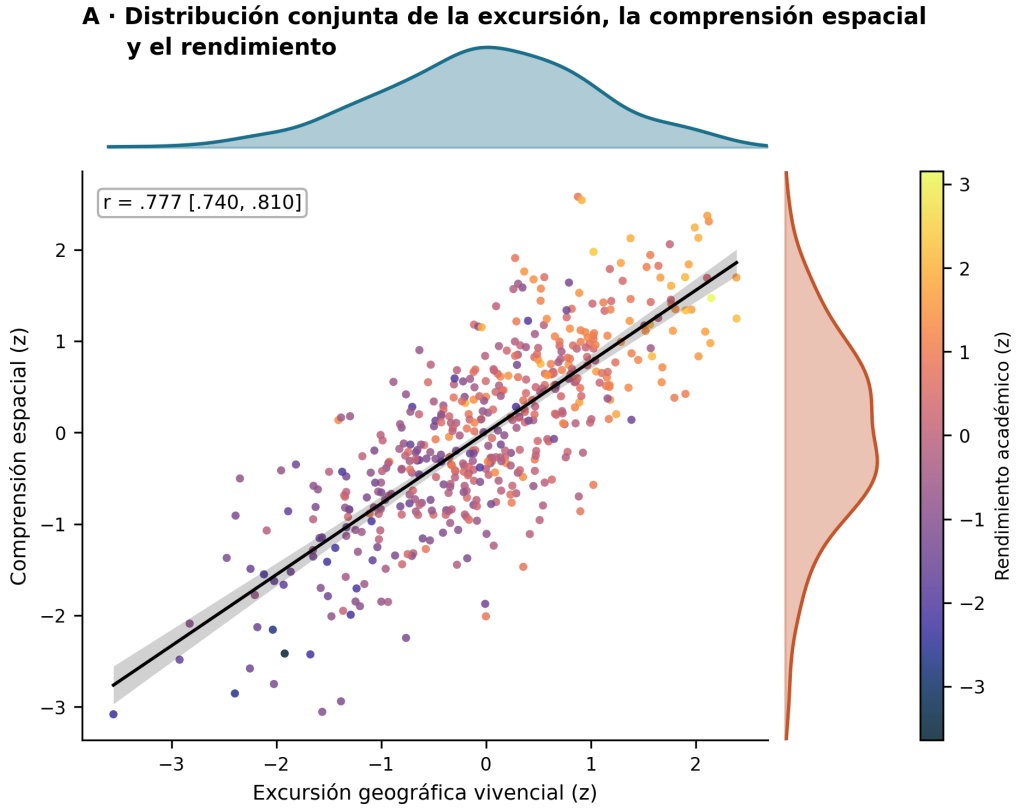
Estadísticos descriptivos y correlaciones entre constructos

Constructo	M	DT	Asimetría	Curtosis	1	2	3
1. Excursión geográfica vivencial	3.001	0.664	−0.214	0.024	—		
2. Comprensión espacial	3.000	0.666	−0.211	0.100	.777*** [.740, .810]	—	
3. Rendimiento académico en geografía	3.001	0.663	−0.090	−0.060	.660*** [.607, .707]	.600*** [.541, .654]	—

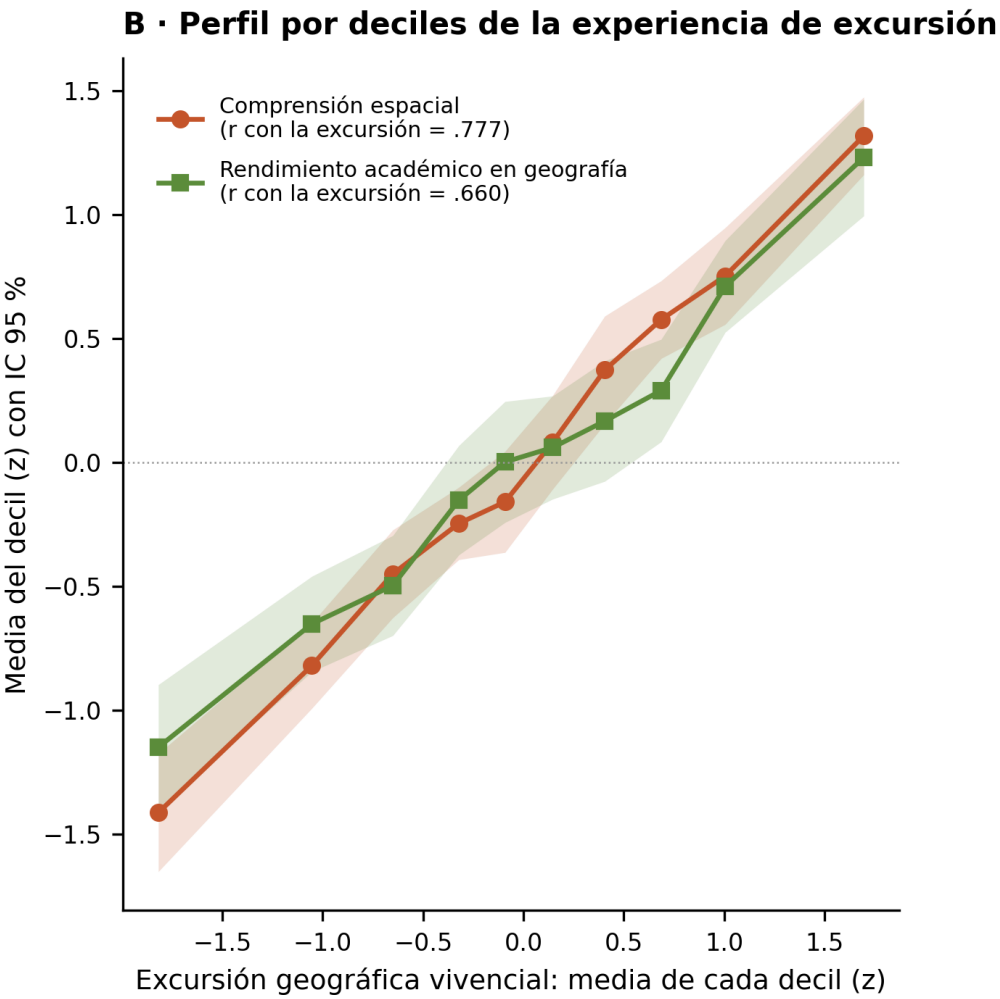
Nota: N = 492. Medias y desviaciones típicas en la escala de respuesta de 1 a 5. Correlaciones de Pearson entre puntuaciones compuestas, con su intervalo de confianza del 95 % (transformación Z de Fisher) entre corchetes. ***p < .001.

Figura 12-1

Distribución y asociación observadas de la excursión geográfica vivencial, la comprensión espacial y el rendimiento académico en geografía

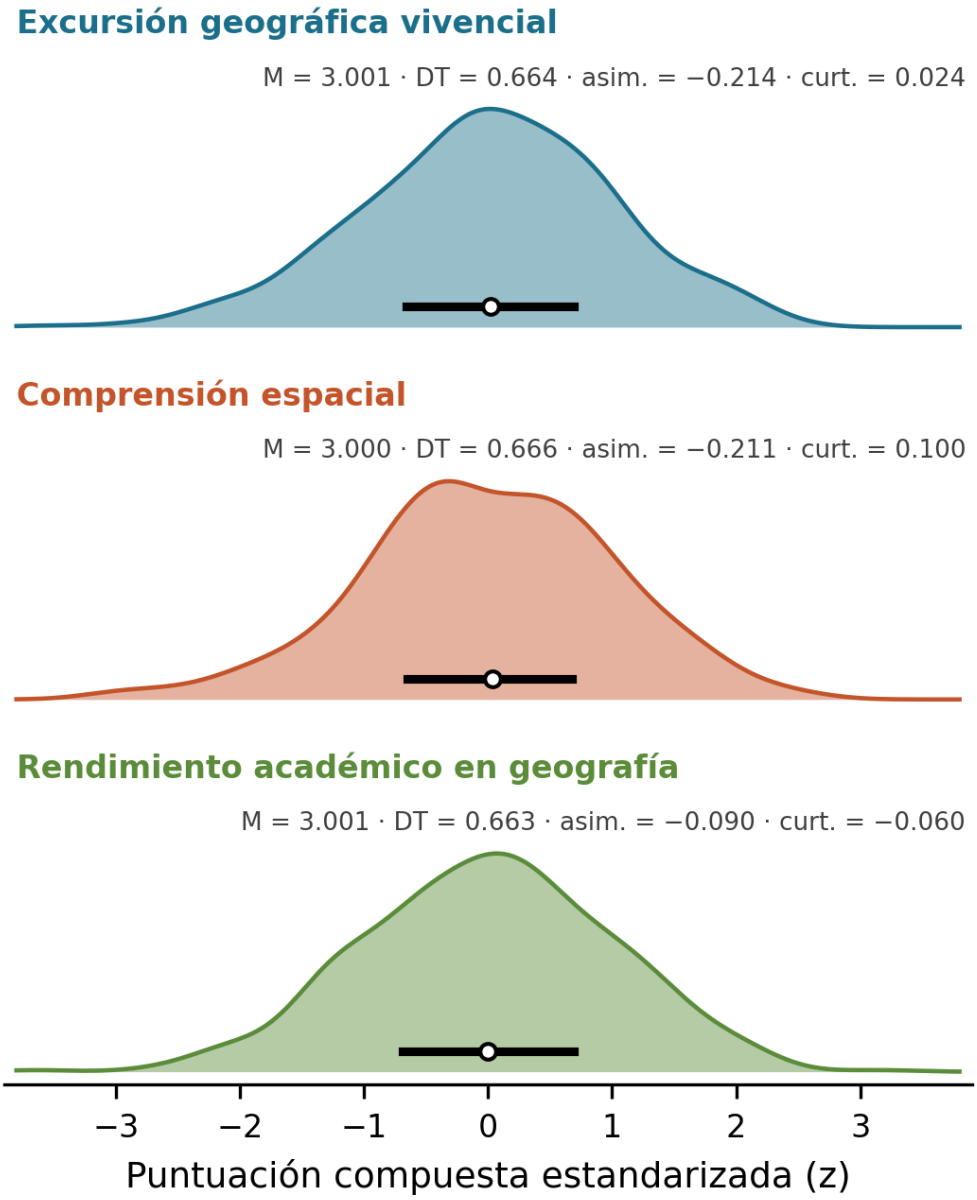


Nota. N = 492; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel A: dispersión de la excursión geográfica vivencial (eje horizontal) y la comprensión espacial (eje vertical), con cada estudiante coloreado según su rendimiento académico en geografía; recta de mínimos cuadrados con banda de confianza del 95 % y densidades marginales de ambos constructos; r = .777 [.740, .810].



Nota. N = 492; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel B: media de la comprensión espacial y del rendimiento académico en geografía en cada decil de la excursión geográfica vivencial, con su intervalo de confianza del 95 %; correlaciones con la excursión de .777 [.740, .810] y .660 [.607, .707].

C · Distribución de los tres constructos



Nota. N = 492; puntuaciones compuestas estandarizadas (z) de las tres escalas. Panel C: densidad de cada constructo con la mediana (punto) y el rango intercuartílico (barra) y sus estadísticos descriptivos de la Tabla 12-3 (excursión geográfica vivencial: M = 3.001, DT = 0.664, asimetría = −0.214, curtosis = 0.024; comprensión espacial: M = 3.000, DT = 0.666, asimetría = −0.211, curtosis = 0.100; rendimiento académico en geografía: M = 3.001, DT = 0.663, asimetría = −0.090, curtosis = −0.060).

12.4.2 Modelo de medida

Todas las cargas estandarizadas superaron .70 y las fiabilidades α y ω superaron .70 en los tres constructos (Tabla 12-2; Figura 12-2A). La validez convergente fue adecuada en los tres, con AVE entre .535 y .661. La validez discriminante se resume en la Tabla 12-4 y la Figura 12-2B: los tres valores de HTMT fueron inferiores a .85, con un máximo de .774 entre la excursión y la comprensión espacial. El criterio de Fornell y Larcker se cumplió en la comprensión espacial y en el rendimiento, pero no en la experiencia de excursión, cuya √AVE (.731) fue menor que su correlación con la comprensión espacial (.777).

Tabla 12-4

Validez discriminante: razón heterorrasgo-monorrasgo y criterio de Fornell-Larcker

Constructo	√AVE	Correlación máxima	Fornell-Larcker	HTMT con		
				excursión	comprensión espacial	rendimiento
Excursión geográfica vivencial	.731	.777	No se cumple	—	.774	.665
Comprensión espacial	.813	.777	Se cumple	.774	—	.645
Rendimiento académico en geografía	.794	.660	Se cumple	.665	.645	—

Nota: N = 492. √AVE = raíz cuadrada de la varianza media extraída; correlación máxima = mayor correlación del constructo con los demás (Tabla 3). Fornell y Larcker: se cumple si √AVE supera la correlación máxima (Fornell y Larcker, 1981). HTMT = razón heterorrasgo-monorrasgo; umbral de .85 (Henseler et al., 2015). Los tres valores HTMT son inferiores a .85.

12.4.3 Ajuste del modelo estructural

El modelo hipotetizado se ajustó a los datos en todos los índices (Tabla 12-5; Figura 12-2C). El modelo alternativo de un solo factor se ajustó claramente peor, con una diferencia de χ^2 significativa ($\Delta\chi^2(3) = 371.069$, $p < .001$) y una pérdida de .172 en el CFI.

Tabla 12-5

Índices de ajuste del modelo hipotetizado y del modelo alternativo de un factor

Modelo	χ^2	gl	p	CFI	TLI	RMSEA [IC 90 %]	p de ajuste cercano	SRMR
Modelo hipotetizado (tres constructos)	16.831	24	.856	1.000	1.000	.000 [.000, .020]	1.000	.016
Modelo alternativo (un factor)	387.900	27	< .001	.828	.771	.165 [.151, .180]	< .001	.107
Diferencia (alternativo – hipotetizado)	$\Delta\chi^2$ = 371.069	Δgl = 3	< .001	ΔCFI = .172	—	—	—	—

Nota: N = 492. Estimación por máxima verosimilitud. El intervalo del RMSEA y la probabilidad de ajuste cercano (H_0: RMSEA ≤ .05) se calcularon con la distribución χ^2 no central. Valores de referencia: CFI y TLI próximos a .95, RMSEA próximo a .06 y SRMR próximo a .08 (Hu y Bentler, 1999).

12.4.4 Modelo estructural y contraste de hipótesis

La Tabla 12-6 presenta las rutas estructurales y la Figura 12-3A, el modelo estimado. Los tres intervalos excluyeron el cero, de modo que H1, H2 y H3 recibieron apoyo.

Tabla 12-6

Coeficientes estructurales del modelo

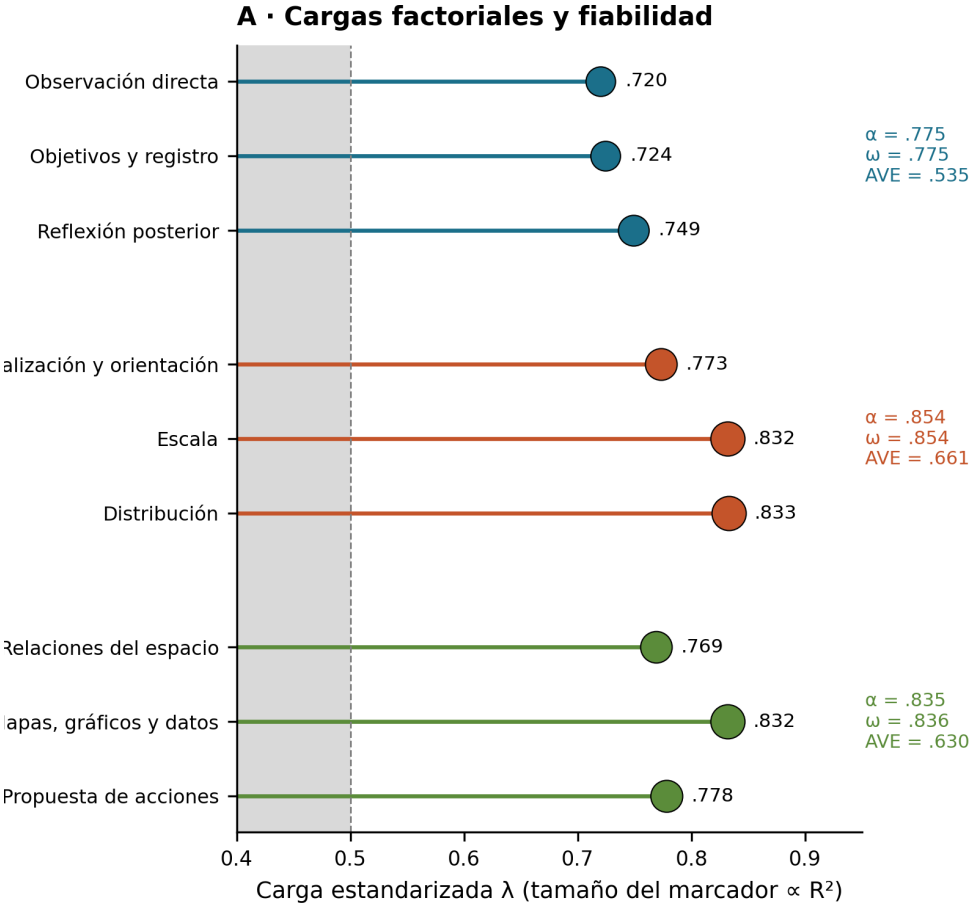
Hipótesis	Ruta	β	p	IC 95 % bootstrap	Decisión
H1	Excursión geográfica vivencial → Rendimiento académico en geografía (c′)	.376	< .001	[.198, .568]	Apoyada
H2	Excursión geográfica vivencial → Comprensión espacial (a)	.751	< .001	[.683, .815]	Apoyada
H3	Comprensión espacial → Rendimiento académico en geografía (b)	.316	< .001	[.114, .499]	Apoyada

Nota: N = 492. β = coeficiente estandarizado del modelo estructural latente estimado por máxima verosimilitud; p = prueba z del coeficiente no estandarizado de la misma ruta. Intervalos de confianza bootstrap percentil del 95 % con 5000 réplicas.

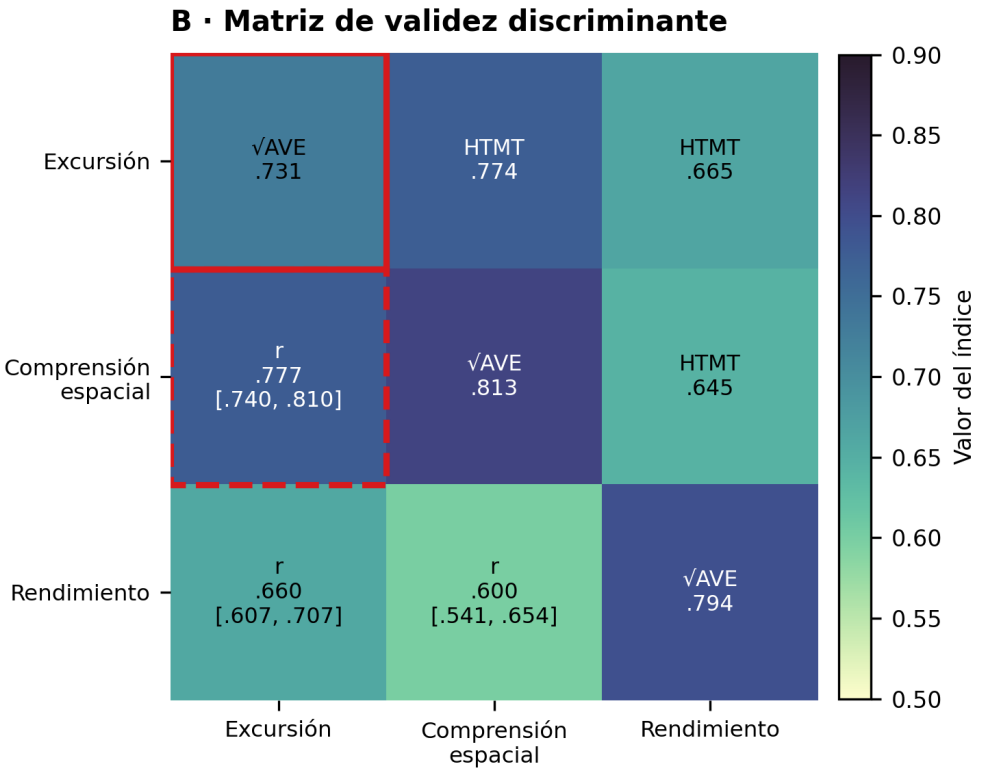
La Tabla 12-7 y la Figura 12-3B descomponen el efecto de la excursión sobre el rendimiento. El efecto indirecto a través de la comprensión espacial fue positivo y su intervalo excluyó el cero, de modo que H4 recibió apoyo.

Figura 12-2

Validez del modelo de medida y ajuste de los modelos competidores

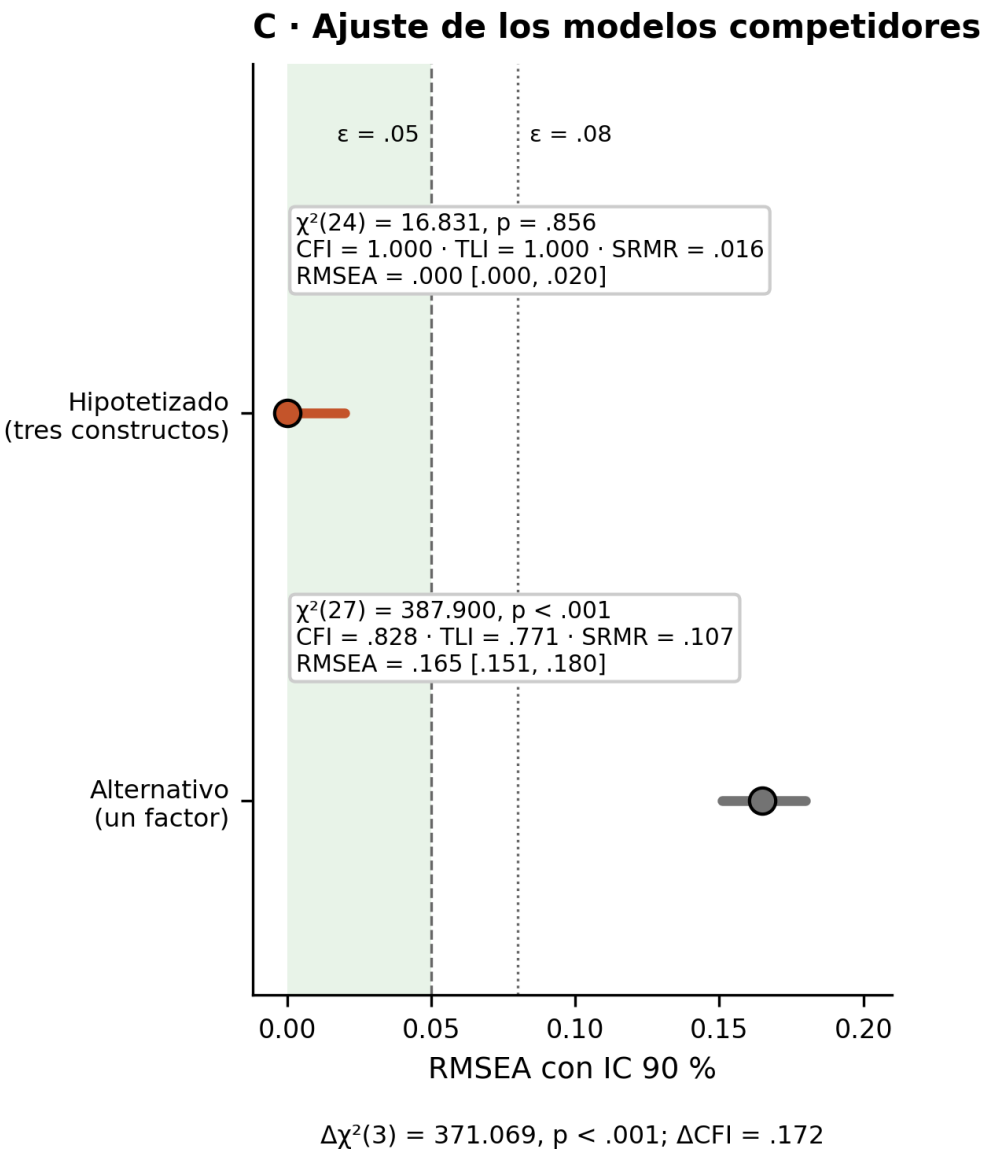


Nota. N = 492. Panel A: cargas factoriales estandarizadas de los nueve indicadores, con el tamaño del marcador proporcional a la varianza explicada de cada ítem y la zona sombreada por debajo de .50; excursión geográfica vivencial: λ = .720–.749, α = .775, ω = .775, AVE = .535; comprensión espacial: λ = .773–.833, α = .854, ω = .854, AVE = .661; rendimiento académico en geografía: λ = .769–.832, α = .835, ω = .836, AVE = .630.



Triángulo inferior: r (Tabla 3) · superior: HTMT (umbral .85) · diagonal: √AVE
Recuadro rojo: √AVE de la excursión (.731) < r con la comprensión espacial (.777)

Nota. N = 492. Panel B: matriz de validez discriminante con la raíz cuadrada de la AVE en la diagonal (.731, .813 y .794), las correlaciones de la Tabla 12-3 con su IC 95 % en el triángulo inferior y los índices HTMT en el superior (.774, .665 y .645; umbral .85); el recuadro rojo señala que la raíz de la AVE de la excursión (.731) no supera su correlación con la comprensión espacial (.777).



Nota. N = 492. Panel C: RMSEA con IC 90 % de los dos modelos frente a los umbrales de .05 y .08; modelo hipotetizado de tres constructos: $\chi^2(24) = 16.831$, $p = .856$; CFI = 1.000; TLI = 1.000; SRMR = .016; RMSEA = .000, IC 90 % [.000, .020]; modelo alternativo de un factor: $\chi^2(27) = 387.900$, $p < .001$; CFI = .828; TLI = .771; SRMR = .107; RMSEA = .165, IC 90 % [.151, .180]; $\Delta\chi^2(3) = 371.069$, $p < .001$; ΔCFI = .172.

El efecto directo siguió siendo distinto de cero junto al indirecto, lo que configura una mediación parcial en la que el 38.7 % del efecto total se transmitió por la comprensión espacial.

Tabla 12-7

Descomposición del efecto de la experiencia de excursión sobre el rendimiento académico

Efecto	Estimación	IC 95 % bootstrap	Distinto de cero
Ruta a: excursión → comprensión espacial	.751	[.683, .815]	Sí
Ruta b: comprensión espacial → rendimiento	.316	[.114, .499]	Sí
Efecto directo c′: excursión → rendimiento	.376	[.198, .568]	Sí
Efecto indirecto a·b (H4)	.237	[.088, .379]	Sí
Efecto total c = c′ + a·b	.613	[.529, .689]	Sí
Proporción indirecta del efecto total (a·b / c)	.387	—	—

Nota: N = 492. Coeficientes estandarizados del modelo estructural latente. Intervalos de confianza bootstrap percentil del 95 % con 5000 réplicas (5000 réplicas convergentes). Un efecto es distinto de cero si su intervalo no incluye el cero.

Como muestra la Tabla 12-8 (Figura 12-4C), el modelo explicó el 56.4 % de la varianza de la comprensión espacial y el 41.9 % de la del rendimiento. En la ecuación del rendimiento, la contribución propia de la excursión (f^2 = .106) fue mayor que la de la comprensión espacial (f^2 = .075).

Tabla 12-8

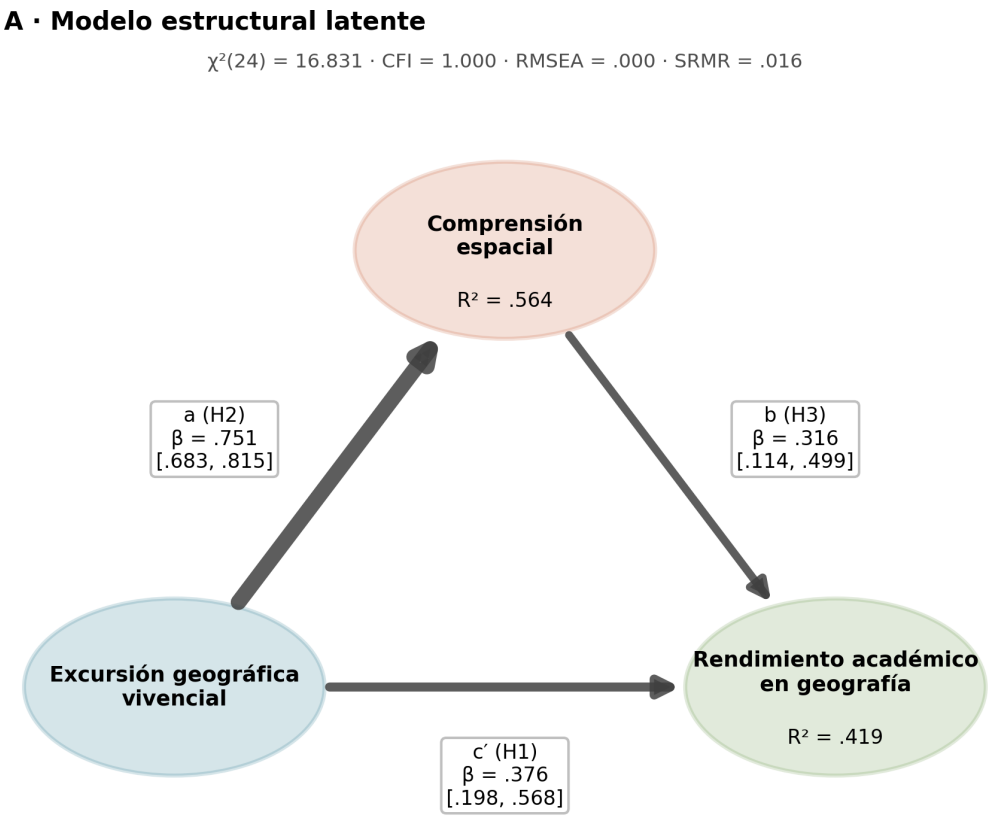
Varianza explicada y tamaño del efecto por ruta

Ecuación	Predictor	β	R^2		f^2
			de la ecuación	sin el predictor	
Comprensión espacial	Excursión geográfica vivencial (a)	.751	.564	.000	1.294
Rendimiento académico en geografía	Excursión geográfica vivencial (c′)	.376	.419	.357	.106
	Comprensión espacial (b)	.316	.419	.375	.075

Nota: N = 492. Valores derivados de los coeficientes estandarizados del modelo estructural: R^2 de la comprensión espacial = a^2; R^2 del rendimiento = $c'^2 + b^2 + 2{\cdot}c'{\cdot}b{\cdot}a$; el R^2 sin un predictor es el cuadrado de la correlación implicada del rendimiento con el otro predictor ($r = c' + a{\cdot}b = .613$ sin la comprensión espacial; $r = b + c'{\cdot}a = .598$ sin la excursión). $f^2 = (R^2_{\text{incluido}} - R^2_{\text{excluido}})/(1 - R^2_{\text{incluido}})$.

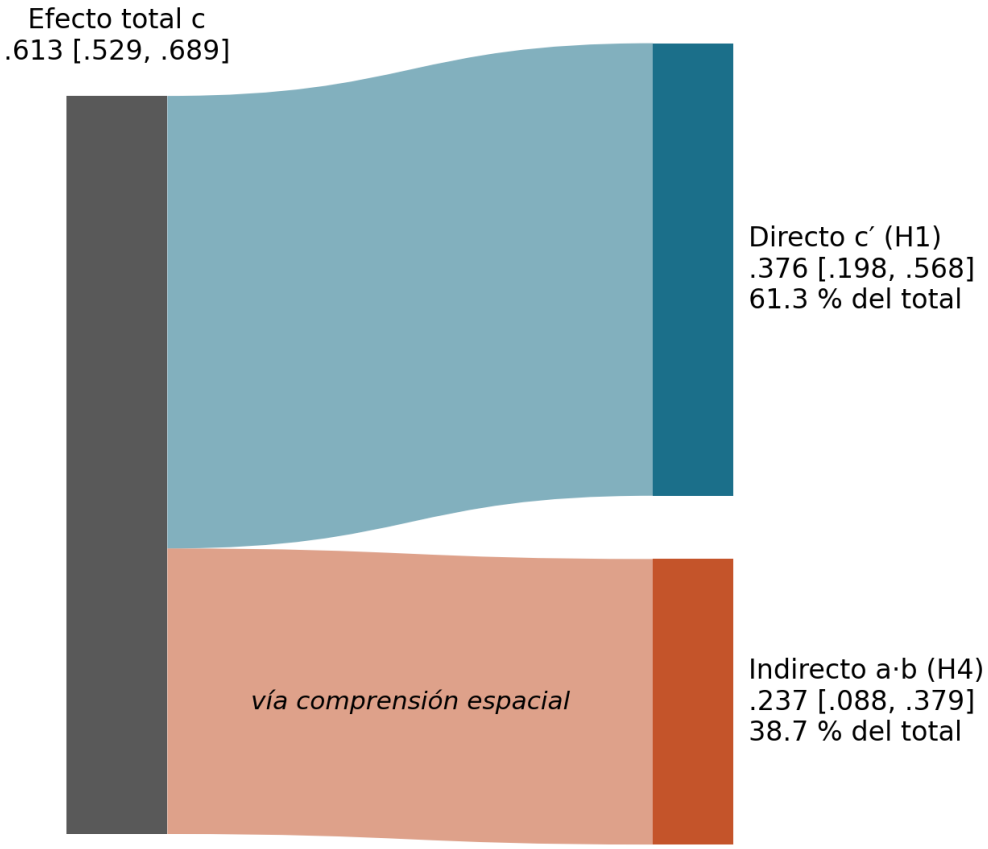
Figura 12-3

Modelo estructural y descomposición del efecto de la excursión geográfica vivencial sobre el rendimiento académico en geografía

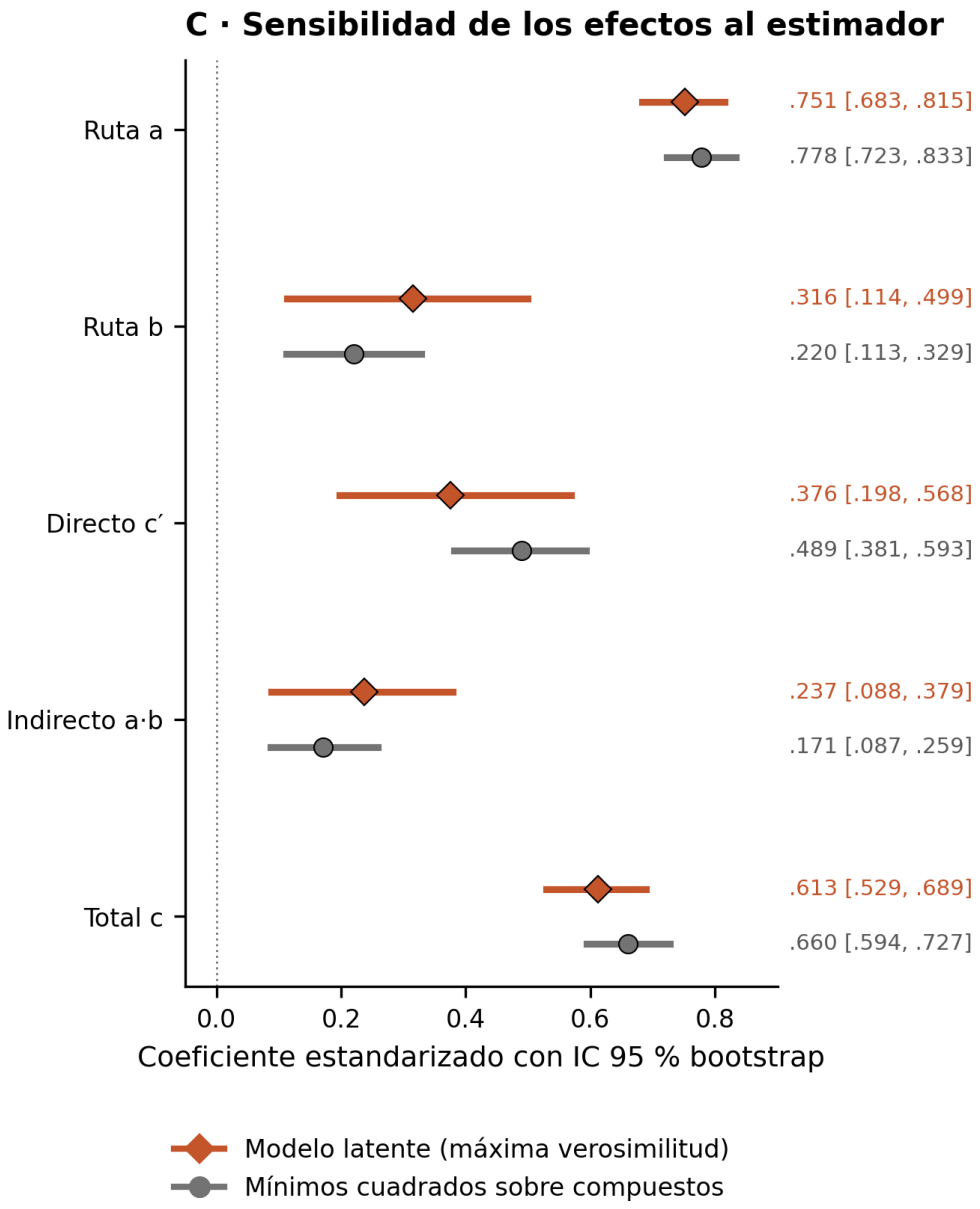


Nota. N = 492; coeficientes estandarizados del modelo de ecuaciones estructurales con estimación de máxima verosimilitud e intervalos de confianza del 95 % por bootstrap percentil (5000 remuestras). Panel A: modelo estructural latente; ruta a = .751 [.683, .815]; ruta b = .316 [.114, .499]; efecto directo c′ = .376 [.198, .568]; R^2 = .564 para la comprensión espacial y .419 para el rendimiento académico en geografía; $\chi^2(24)$ = 16.831, CFI = 1.000, RMSEA = .000, SRMR = .016; el grosor de cada flecha es proporcional a su coeficiente.

B · Descomposición del efecto total



Nota. N = 492; coeficientes estandarizados del modelo de ecuaciones estructurales con estimación de máxima verosimilitud e intervalos de confianza del 95 % por bootstrap percentil (5000 remuestras). Panel B: descomposición del efecto total c = .613 [.529, .689] en el efecto directo (.376 [.198, .568]; 61.3 %) y el efecto indirecto a través de la comprensión espacial (.237 [.088, .379]; 38.7 %).



Nota. N = 492; coeficientes estandarizados del modelo de ecuaciones estructurales con estimación de máxima verosimilitud e intervalos de confianza del 95 % por bootstrap percentil (5000 remuestras). Panel C: comparación de cada efecto estimado con el modelo latente (rombos) y con mínimos cuadrados ordinarios sobre las puntuaciones compuestas (círculos; IC 95 % por bootstrap percentil de 5000 remuestras): ruta a, .751 [.683, .815] frente a .778 [.723, .833]; ruta b, .316 [.114, .499] frente a .220 [.113, .329]; efecto directo c′, .376 [.198, .568] frente a .489 [.381, .593]; efecto indirecto a·b, .237 [.088, .379] frente a .171 [.087, .259]; efecto total c, .613 [.529, .689] frente a .660 [.594, .727].

12.4.6 Sensibilidad al estimador y estabilidad del efecto

La re-estimación por mínimos cuadrados sobre puntuaciones compuestas reprodujo la dirección y la significación de todos los efectos (Tabla 9; Figura 3C). La ruta a y el efecto directo fueron algo mayores, y la ruta b y el efecto indirecto algo menores que en el modelo latente; el efecto indirecto siguió siendo distinto de cero con ambos estimadores.

Tabla 12-9

Sensibilidad de los efectos al estimador

Efecto		Modelo latente:		Puntuaciones compuestas:	
		máxima verosimilitud	IC 95 %	Mínimos cuadrados	IC 95 %
a:	excursión → comprensión espacial	.751	[.683, .815]	.778	[.723, .833]
b:	comprensión espacial → rendimiento	.316	[.114, .499]	.220	[.113, .329]
c′:	efecto directo	.376	[.198, .568]	.489	[.381, .593]
a × b:	efecto indirecto	.237	[.088, .379]	.171	[.087, .259]
c:	efecto total	.613	[.529, .689]	.660	[.594, .727]

Nota: N = 492. Modelo latente: coeficientes estandarizados con intervalos bootstrap percentil (5000 réplicas). Mínimos cuadrados: regresiones sobre puntuaciones compuestas estandarizadas de los tres constructos, con intervalos bootstrap percentil (5000 réplicas, semilla fija).

El efecto total fue estable ante las perturbaciones de la muestra (Tabla 12-10; Figura 12-4B): la pendiente se mantuvo entre 0.390 y 0.446 en todas las variantes, conservó el signo y la significación en el 100 % de las réplicas de jackknife, submuestreos y bootstrap, y desapareció en la prueba placebo.

Tabla 12-10

Estabilidad del efecto total de la experiencia de excursión sobre el rendimiento académico

Prueba	Réplicas o n	Pendiente	DT	IC 95 %	Signo estable (%)	Significativo (%)
Jackknife (dejar uno fuera)	492 réplicas	0.433	0.001	[0.430, 0.435]	100 %	100 %
Submuestreo del 80 %	100 réplicas	0.431	0.012	[0.404, 0.452]	100 %	100 %
Submuestreo del 60 %	100 réplicas	0.431	0.019	[0.391, 0.463]	100 %	100 %
Bootstrap	1000 réplicas	0.432	0.021	[0.389, 0.470]	100 %	100 %
Rendimiento winsorizado al 5 %	n = 492	0.390	—	[0.355, 0.425]	—	p < .001
Submuestra única del 60 %	n = 295	0.446	—	[0.396, 0.497]	—	p < .001
Placebo (excursión permutada)	n = 492	−0.034	—	[−0.097, 0.030]	—	p = .300

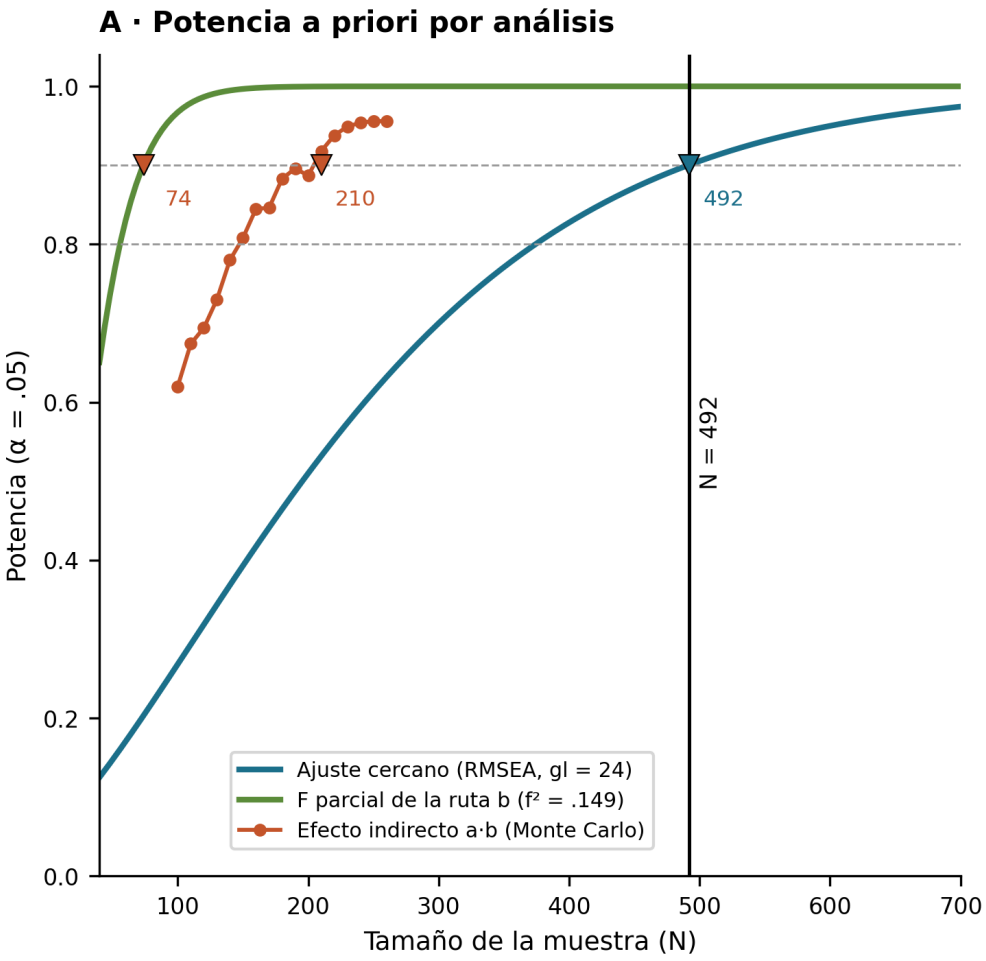
Nota: Pendiente no estandarizada de la regresión por mínimos cuadrados del rendimiento académico sobre la experiencia de excursión (puntuaciones compuestas; pendiente en la muestra completa = 0.433). En las cuatro primeras filas, la pendiente es la media de las réplicas, DT su desviación típica y el IC el intervalo percentil del 95 % de las réplicas; «signo estable» y «significativa» son los porcentajes de réplicas con signo positivo y con p < .05. En las tres últimas filas, el IC es el de la estimación única.

12.5 DISCUSIÓN

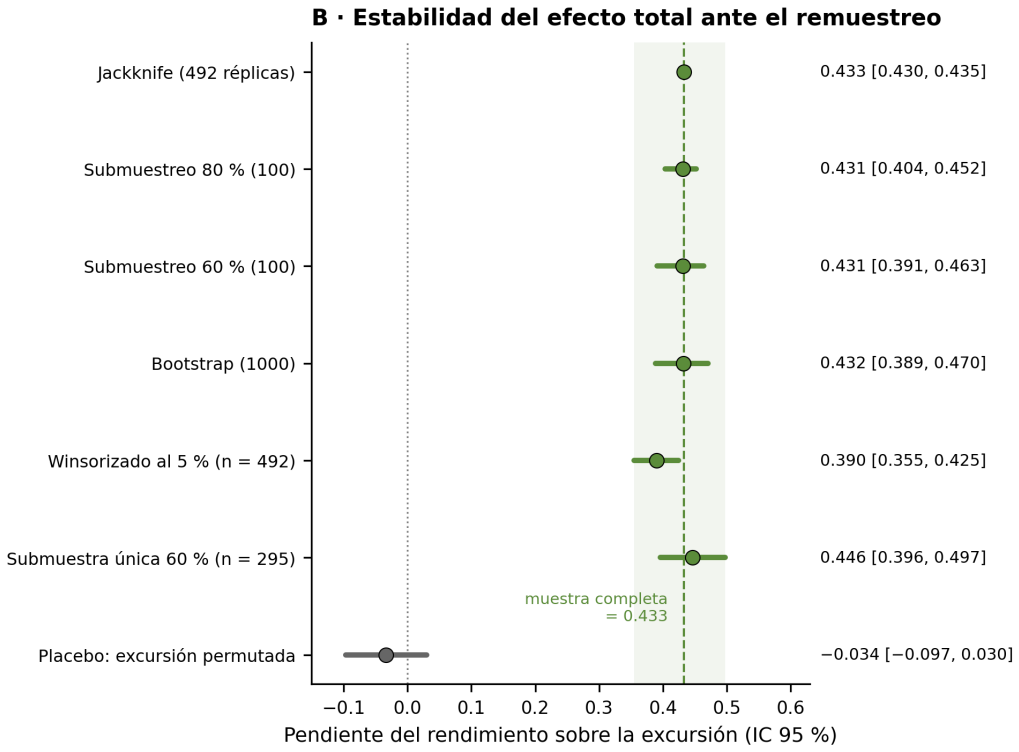
El hallazgo más desconcertante no es que la comprensión espacial medie la relación entre la excursión y el rendimiento, sino que lo haga en una proporción menor de la esperada. La ruta de la excursión hacia la comprensión espacial es, con diferencia, la más intensa del modelo, y sin embargo el 61.3 % del efecto total no transita por ella: el efecto directo conserva un peso propio superior al de la comprensión espacial en la ecuación del rendimiento. Si la salida de campo enseñara geografía principalmente porque enseña a pensar espacialmente, cabría esperar un efecto directo residual mucho más modesto.

Figura 12-4

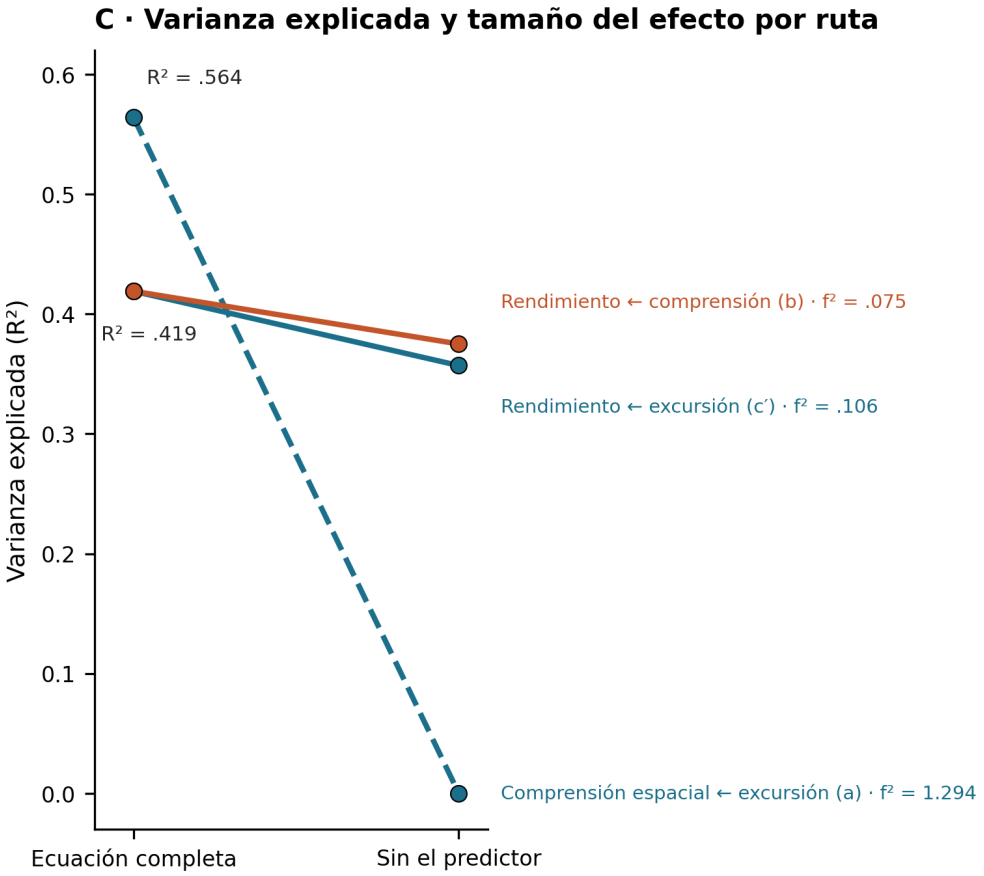
Potencia a priori, estabilidad del efecto y varianza explicada por ruta



Nota. N = 492. Panel A: potencia (α = .05) en función del tamaño de la muestra para la prueba de ajuste cercano del RMSEA (gl = 24; ε0 = .05 frente a εa = .08), la prueba F parcial de la ruta b (f² = .149) y el efecto indirecto a·b por simulación de Monte Carlo (a = .229, b = .36); los triángulos marcan el N mínimo para una potencia de .90 (492, 74 y 210, respectivamente) y la línea vertical, el N analizado (potencia del ajuste cercano = .901).



Nota. N = 492. Panel B: pendiente no estandarizada del rendimiento académico en geografía sobre la excursión geográfica vivencial (muestra completa = 0.433; banda sombreada: rango de los intervalos de las seis réplicas) con IC 95 %: jackknife con 492 réplicas, 0.433 [0.430, 0.435]; submuestreo del 80 % (100 réplicas), 0.431 [0.404, 0.452]; submuestreo del 60 % (100 réplicas), 0.431 [0.391, 0.463]; bootstrap (1000 réplicas), 0.432 [0.389, 0.470]; rendimiento winsorizado al 5 %, 0.390 [0.355, 0.425]; submuestra única del 60 % (n = 295), 0.446 [0.396, 0.497]; placebo con la excursión permutada, –0.034 [–0.097, 0.030].



Nota. N = 492. Panel C: R^2 de la ecuación completa y R^2 al excluir cada predictor, con su f^2: comprensión espacial ← excursión (a), R^2 = .564 frente a .000, f^2 = 1.294; rendimiento ← excursión (c′), R^2 = .419 frente a .357, f^2 = .106; rendimiento ← comprensión espacial (b), R^2 = .419 frente a .375, f^2 = .075.

Los datos sugieren, en cambio, una mediación parcial en la que el mecanismo espacial explica una parte sustantiva, pero no mayoritaria, de la asociación, lo que respalda las cuatro hipótesis y obliga a preguntarse qué más contiene una excursión bien planificada.

La literatura sobre aprendizaje experiencial ofrece candidatos que el modelo no mide y que, desde una perspectiva crítica, conviene no tratar como residuo estadístico. Los entornos experienciales eficaces se caracterizan por la autenticidad de la tarea, una estructura explícita y la agencia del estudiante, rasgos asociados con resultados cognitivos, afectivos, metacognitivos y socio-comunicativos (Duchatelet et al., 2024); es verosímil que parte del efecto directo refleje esa implicación y no una ganancia espacial. En la misma dirección, los docentes y estudiantes consultados en un estudio sobre salidas de campo atribuyen a estas un impacto percibido más elevado sobre el bienestar social y personal que sobre el avance académico (Lynch et al., 2026), de modo que un componente motivacional o relacional podría estar contribuyendo al rendimiento autopercibido sin pasar por el razonamiento espacial. Otros trabajos plantean que las salidas culturalmente pertinentes pueden dejar de ser actividades de enriquecimiento aisladas y convertirse en experiencias transformadoras que mejoran los resultados académicos (Fish, 2026), y la investigación clásica ya señalaba que la eficacia de una salida depende de su calidad y de la familiaridad previa con el entorno (Orion y Hofstein, 1994). Esta lectura es, no obstante, conjetural: el diseño no permite aislar esos mecanismos, y una parte del efecto directo podría deberse también a varianza común de método, como se discute más abajo. La mediación parcial es, por ello, un resultado informativo, porque delimita el alcance del mecanismo espacial y señala qué mediadores —motivación, interés por la disciplina, vínculo con el lugar— deberían incorporarse a modelos posteriores.

Las rutas a y b, consideradas por separado, son consistentes con lo que cabía anticipar. La intensidad de la ruta a concuerda con la evidencia de que la experiencia y el entrenamiento mejoran las habilidades espaciales (Uttal et al., 2013) y con los resultados de una propuesta de campo en la

que el entrenamiento con mapas mentales y realidad virtual consolidó los marcos cognitivos necesarios para la navegación real (Czomba et al., 2026). La ruta b, más moderada, apunta en la misma dirección que la asociación entre habilidades espaciales y rendimiento matemático (Atit et al., 2022) y que las mejoras de pensamiento espacial, rendimiento y actitudes tras una intervención con tecnologías espaciales en primaria (Özdemir, 2026).

La implicación didáctica trasciende el caso peruano. Los ítems de la excursión miden su componente planificado —observación directa, objetivos y registro, reflexión posterior—, no la mera asistencia, de modo que, si la asociación con el rendimiento pasa en parte por la comprensión espacial, la prioridad al diseñar una salida sería acompañar la observación con tareas de orientación, escala y distribución, y conectar lo observado con su representación. Esta consideración es pertinente para el debate sobre la sustitución parcial de las salidas por recorridos virtuales, que ganaron atención durante la pandemia por su capacidad de trascender el tiempo y el espacio (Hu y Zhang, 2026) y cuyo desarrollo y dotación de recursos constituyen un reto específico (Hurrell et al., 2025). Los datos de este estudio no comparan formatos, pero sugieren que lo que debe preservarse, sea cual sea el soporte, es lo que entrena el razonamiento espacial. Las experiencias inmersivas podrían sostener parte de ese efecto, dado que la inmersión en realidad virtual se ha asociado con mayor retención y presencia que el vídeo (Makransky y Mayer, 2022), y el trabajo con realidad virtual en adolescentes subraya la necesidad de metas de aprendizaje explícitas y de práctica regular de trabajo de campo local (Palsdottir, 2025). A ello se suma una salvedad: el aprovechamiento de los entornos virtuales puede depender de la propia habilidad espacial del estudiante (Guo et al., 2025), lo que implicaría riesgos de desigualdad si el formato virtual reemplazara la experiencia directa. En sistemas escolares donde el trabajo de campo está previsto en el currículo pero no siempre se ejecuta, como se ha documentado con docentes de geografía en Indonesia (Oktavianto et al., 2024a), el modelo ofrece un criterio para decidir cómo emplear un tiempo de campo escaso: concentrarlo en tareas espaciales explícitas.

La robustez de los resultados descarta varias lecturas alternativas. El modelo de tres constructos superó con claridad al de un solo factor, de modo que las tres medidas no son indistinguibles. La re-estimación sobre puntuaciones compuestas reprodujo la dirección y la significación de todos los efectos, y las diferencias de magnitud siguen el patrón esperable de un estimador que no corrige el error de medida: sobreestima la ruta a y el efecto directo e infraestima la ruta b y el indirecto. El efecto total resistió el jackknife, los submuestreos y el bootstrap, y se desvaneció en la prueba placebo, lo que hace poco probable que dependa de observaciones influyentes o de un artefacto de estimación.

Esa robustez no neutraliza las limitaciones del diseño. En primer lugar, la naturaleza transversal impide afirmar causalidad: las rutas son asociaciones compatibles con el modelo propuesto, y la dirección inversa —que los estudiantes con mayor comprensión espacial aprovechen más la excursión— no puede descartarse. En segundo lugar, los tres constructos se midieron por autoinforme en un mismo cuestionario, lo que puede inflar las correlaciones por varianza común de método, y el rendimiento se midió como logro percibido de la competencia, no con calificaciones ni pruebas estandarizadas. En tercer lugar, la validez discriminante entre la excursión y la comprensión espacial descansa en el HTMT, que detecta con más fiabilidad la falta de discriminación que el criterio de Fornell y Larcker (Henseler et al., 2015), pero no en este último (Fornell y Larcker, 1981); ambas medidas comparten una proporción alta de varianza y la escala de excursión, con una AVE de .535, debería ampliarse y refinarse. En cuarto lugar, la muestra es no probabilística y por conveniencia, por lo que los resultados describen a los estudiantes analizados. Por último, los estudiantes están agrupados en instituciones y el modelo no estima la variación entre escuelas.

Estas limitaciones definen la agenda siguiente: diseños longitudinales o cuasi-experimentales que comparen grupos con y sin excursión planificada; medidas objetivas del pensamiento espacial, como las pruebas estandarizadas de habilidad espacial (Lee y Bednarz, 2012), y del rendimiento, mediante calificaciones o evaluaciones curriculares;

modelos multinivel que separen la variación entre escuelas; y modelos que incorporen mediadores no espaciales para explicar el efecto directo.

12.6 CONCLUSIONES

La experiencia de excursión geográfica vivencial se asocia con el rendimiento académico en geografía de estudiantes de secundaria del Perú por dos vías: una indirecta, a través de la comprensión espacial, que transmite el 38.7 % del efecto total, y una directa que conserva peso propio. El modelo de mediación parcial se ajustó a los datos, superó con claridad a un modelo de un solo factor y se mostró robusto frente al estimador y a las perturbaciones de la muestra.

Para la teoría del aprendizaje geográfico experiencial, el estudio aporta una contrastación del mecanismo espacial que la literatura suponía, y a la vez muestra sus límites: la comprensión espacial es una vía relevante, pero no la única, y la parte no espacial del efecto reclama modelos que incorporen mediadores motivacionales, de interés por la disciplina o de vínculo con el lugar. Para la práctica docente, el resultado ofrece un criterio concreto: la excursión rinde más cuando su planificación —objetivos, registro y reflexión posterior— se orienta a entrenar la localización, la escala y la distribución. Ese componente es el que conviene preservar cuando la salida se acorta, se reduce o se sustituye por formatos virtuales, y es el que justifica concentrar el escaso tiempo de campo en tareas espaciales explícitas.

Estas conclusiones deben leerse con la cautela que imponen un diseño transversal, mediciones de autoinforme y una muestra por conveniencia: los coeficientes son asociaciones, no efectos causales, y describen a los estudiantes analizados. La validación con diseños longitudinales o cuasi-experimentales, con medidas objetivas de pensamiento espacial y de rendimiento y con modelos multinivel, es la tarea pendiente para establecer hasta dónde se sostiene este modelo en otros sistemas escolares.

12.7 REFERENCIAS BIBLIOGRÁFICAS

Alazmi, H., & Alemtairy, G. (2024). The effects of immersive virtual reality field trips upon student academic achievement, cognitive load, and multimodal presence in a social studies educational context. *Education and Information Technologies, 29*(16), 22189-22211. https://doi.org/10.1007/s10639-024-12682-3

Andreis, A., & Claudino, S. (2024). La construcción del pensamiento geográfico en la enseñanza escolar. *Boletín de Estudios Geográficos*(122), 211-230. https://doi.org/10.48162/rev.40.056

Araya, F., & Álvarez, S. (2025). Cartografia escolar para o desenvolvimento do pensamento geográfico: diálogos e perspectivas entre Brasil e Chile. *Ateliê Geográfico, 19*(1), 117-143. https://doi.org/10.5216/ag.v19i1.81388

Araya, F., & Álvarez, S. (2026). Educación geográfica desde una perspectiva post-humanista. Desafíos para el desarrollo del pensamiento espacial. *Revista de geografía Norte Grande*(94), 1-22. https://doi.org/10.4067/s0718-34022026000200104

Atit, K., Power, J., Pigott, T., Lee, J., Geer, E., Uttal, D., Ganley, C., & Sorby, S. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review, 29*(3), 699-720. https://doi.org/10.3758/s13423-021-02012-w

Aylward, T., Coulter, M., & Farrelly, T. (2026). Outdoor learning on the edge of Europe: a systematic review of practice from Ireland. *Journal*

of Adventure Education and Outdoor Learning, *26*(3), 553-576. https://doi.org/10.1080/14729679.2025.2494033

Cao, K., Qi, Y., Neo, H., & Guo, H. (2024). Web GIS as a pedagogical tool in tourist geography course: the effect on spatial thinking ability and self-efficacy. *Journal of Geography in Higher Education*, *48*(2), 193-210. https://doi.org/10.1080/03098265.2023.2198200

Cascante, A. (2024). El valor de la geografía en la educación: el conocimiento poderoso en geografía (powerful geography) y la educación geográfica. *Papeles de Geografía*(69). https://doi.org/10.6018/geografia.568281

Chocce, F. (2026). Desempeño docente y logros de aprendizaje en estudiantes de secundaria rural del Perú. *Llimpi*, *6*(1), 51-56. https://doi.org/10.54943/lree.v6i1.811

Crane, A., & Mewborne, M. (2026). On Spatial Thinking and Spatial Citizenship in Geography Education: Lessons from Editing a State Atlas. *The Geography Teacher*, *23*(3), 168-175. https://doi.org/10.1080/19338341.2026.2673095

Czomba, P., Czimre, K., Teperics, K., Bujdosó, G., Molnár, E., Négyesi, G., & Juhász, B. (2026). Enhancing Spatial Orientation and Map-Reading Skills: Using Mental Maps and VR in Field Trips for Geography Students. *ISPRS International Journal of Geo-Information*, *15*(5), 227. https://doi.org/10.3390/ijgi15050227

De Sousa, P., & Barbosa, E. (2025). O trabalho de campo no ensino de Geografia: articulação entre os componentes físico-naturais e o uso

das tecnologias da informação e comunicação. *Para Onde!?*, *19*(2). https://doi.org/10.22456/1982-0003.136270

De Sousa, S., & Sepúlveda, O. (2024). Trabajo de campo en la enseñanza de Geografía e Historia: experiencias formativas en Brasil y Chile. *Ateliê Geográfico*, *18*(1), 400-428. https://doi.org/10.5216/ag.v18i1.79021

Dobečková, M., & Maršálek, R. (2026). Learning Without Walls: A Systematic Review of Evaluation Approaches and Educational Outcomes in Chemistry and Science Field Trip Research. *Education Sciences*, *16*(9), 1378. https://doi.org/10.3390/educsci16091378

Duchatelet, D., Cornelissen, F., & Volman, M. (2024). Features of Experiential Learning Environments in Relation to Generic Learning Outcomes in Higher Education: A Scoping Review. *Journal of Experiential Education*, *47*(3), 400-423. https://doi.org/10.1177/10538259231211537

Díaz, M., & Galán, V. (2026). La enseñanza de la geografía y la educación ambiental en educación secundaria. *Revista RedCA*, *8*(24), 132. https://doi.org/10.36677/redca.v8i24.27526

Evangelista, V., Menezes, I., Nascimento, T., & Silva, M. (2026). Entre teoria e prática: do pré-campo ao trabalho de campo no ensino de Geografia. *Para Onde!?*, *20*(1). https://doi.org/10.22456/1982-0003.150895

Fish, L. (2026). Curating culturally responsive field trips: toward continuous and transformative experiential learning. *Social Studies Research and Practice*, *21*(1), 14-25. https://doi.org/10.1108/ssrp-06-2025-0029

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Guallart, C. (2025). Competencia espacial, pensamiento geográfico y ciudadanía crítica: un enfoque integrado en educación secundaria. *Espacio Tiempo y Forma. Serie VI, Geografía*(18), 151-180. https://doi.org/10.5944/etfvi.18.2025.45337

Guo, S., Sun, Y., & Wang, X. (2025). Does spatial ability affect students' learning performance in virtual reality environments? A study on desktop virtual reality systems. *Thinking Skills and Creativity, 58,* 101923. https://doi.org/10.1016/j.tsc.2025.101923

Gómez, A., Font, D., Sastre, B., & Binimelis, J. (2025). ¿Qué geografía de España conocen los alumnos de Educación Secundaria Obligatoria? *Didáctica Geográfica*(26), 241-262. https://doi.org/10.21138/dg.749

Haquet, E., & Aguilera, L. (2025). La enseñanza de la geografía con un enfoque experimental. *Tlalli. Revista de Investigación en Geografía*(13), 107-137. https://doi.org/10.22201/ffyl.26832275e.2025.13.2191

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.

Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1-55. https://doi.org/10.1080/10705519909540118

Hu, R., & Zhang, Y. (2026). A New Approach to Field Learning—Virtual Field Trips: A Systematic Review. *Journal of Computer Assisted Learning, 42*(1). https://doi.org/10.1002/jcal.70170

Hurrell, E., Hutchinson, S., Yorke, L., Batty, L., Bunting, M., Swanton, D., McDougall, D., & Parsons, D. (2025). The role of virtual field trips in Geography higher education: A perspective paper. *Area, 57*(3). https://doi.org/10.1111/area.70011

Lee, J., & Bednarz, R. (2012). Components of Spatial Thinking: Evidence from a Spatial Thinking Ability Test. *Journal of Geography, 111*(1), 15-26. https://doi.org/10.1080/00221341.2011.583262

Lee, J., Cimová, T., Foster, E., France, D., Krajňáková, L., Moorman, L., Rewhorn, S., & Zhang, J. (2025). Transforming geography education: the role of generative AI in curriculum, pedagogy, assessment, and fieldwork. *International Research in Geographical and Environmental Education, 34*(3), 237-253. https://doi.org/10.1080/10382046.2025.2459780

Liang, X., Dong, D., Zhang, J., & Ge, J. (2026). Beyond the entity perspective: Exploring the internal development mechanism of high school students' geospatial thinking based on network analysis. *Thinking Skills and Creativity, 61*, 102188. https://doi.org/10.1016/j.tsc.2026.102188

Lynch, R., McCormack, O., & Hennessy, J. (2026). Exploring teachers' and students' views and experiences of field trips. *Journal of Adventure*

Education and Outdoor Learning, *26*(3), 702-720. https://doi.org/10.1080/14729679.2025.2530091

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Makransky, G., & Mayer, R. (2022). Benefits of Taking a Virtual Field Trip in Immersive Virtual Reality: Evidence for the Immersion Principle in Multimedia Learning. *Educational Psychology Review*, *34*(3), 1771-1798. https://doi.org/10.1007/s10648-022-09675-4

Mantilla, L., Zambrano, A., Cuenca, M., Quinga, J., & Gallegos, E. (2025). El Aporte de las Salidas Pedagógicas al Proceso de Enseñanza-Aprendizaje. *Ciencia Latina Revista Científica Multidisciplinar*, *9*(2), 1340-1354. https://doi.org/10.37811/cl_rcm.v9i2.16976

Megoran, N., & Benwell, M. (2026). Teaching hopeful political geographies through field trips. *Journal of Geography in Higher Education*, 1-12. https://doi.org/10.1080/03098265.2026.2699177

Menezes, I. (2025). Planejando trabalho de campo: encontros entre a Geografia acadêmica e escolar a partir da análise da paisagem sob a perspectiva socioambiental em Ilhéus-BA. *Revista de Ensino de Geografia*, *16*(30). https://doi.org/10.14393/reg-v16-2025-80896

Ministerio de Educación del Perú. (2016). *Currículo Nacional de la Educación Básica.* Minedu.

Minor, J., & McCourt, M. (2021). Role-Based Assignments Support Student Engagement during Geography Field Trips. *The Geography Teacher*, *18*(1), 24-38. https://doi.org/10.1080/19338341.2020.1861054

Moreno, M., Abel, J., & López, Y. (2023). Pensamiento geográfico, método para orientar la educación geográfica en el bachillerato. *Portal de la Ciencia*, *4*(1), 82-99. https://doi.org/10.51247/pdlc.v4i1.339

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Oktavianto, D., Sumarmi, S., Utaya, S., Taryana, D., & Widodo, B. (2024a). Challenges in the implementation of fieldwork learning in the geography curriculum in Indonesia. *Journal of the Geographical Institute Jovan Cvijic, SASA*, *74*(2), 281-289. https://doi.org/10.2298/ijgi2304260060

Oktavianto, D., Utaya, S., Sumarmi, S., & Taryana, D. (2024b). Implementing fieldwork based on blended learning using geo-inquiry approach: will it work during the covid-19 pandemic? *International Research in Geographical and Environmental Education*, *33*(2), 125-138. https://doi.org/10.1080/10382046.2023.2214043

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching*, *31*(10), 1097-1119. https://doi.org/10.1002/tea.3660311005

Özdemir, C. (2026). The impact of spatial information technologies on spatial thinking, academic achievement, and learner attitudes among third-grade primary students. *Children's Geographies*, *24*(5), 557-575. https://doi.org/10.1080/14733285.2026.2704533

Palsdottir, A. (2025). Activating teenagers' learning in geography using virtual reality: student enquiry of local school environment in Iceland. *International Research in Geographical and Environmental Education*, *34*(1), 43-59. https://doi.org/10.1080/10382046.2024.2349427

Pulido, L., Pépin, A., Bergeron, C., Cherblanc, J., Godue, C., Laprise, C., Paquette, L., Nadeau, S., & Simard, S. (2025). The Effects of Outdoor Teaching on Academic Achievement and Its Associated Factors—A Scoping Review. *Education Sciences*, *15*(8), 1060. https://doi.org/10.3390/educsci15081060

Quintero, M. (2026). Abrir la geografía escolar. Estado del arte sobre las propuestas desde la educación geográfica en Colombia (2018-2023). *Revista Geográfica*(169), 55-74. https://doi.org/10.35424/regeo.169.2024.5805

Remmen, K., & Iversen, E. (2023). A scoping review of research on school-based outdoor education in the Nordic countries. *Journal of Adventure Education and Outdoor Learning*, *23*(4), 433-451. https://doi.org/10.1080/14729679.2022.2027796

Retana, D., & Cascante, J. (2026). Significando el trabajo del campo en Geografía. *Revista Brasileira de Educação em Geografia*, *16*(26), 12-37. https://doi.org/10.46789/edugeo.v16i26.1656

Rodriguez, M., & Aguilar, A. (2026). When GIS Succeeds: Fostering Critical Spatial Thinking through Geography Education on Rural Depopulation. *European Journal of Geography*, *17*(2), 302-317. https://doi.org/10.48088/ejg.m.ang.17.2.302.317

Romero, O. (2025). Formación en pensamiento espacial para la enseñanza de la Geografía. *Ar@cne. Revista Electrónica de Recursos en Internet sobre Geografía y Ciencias Sociales, 29*. https://doi.org/10.1344/ara2025.303.49607

Rosero, E., & Gómez, Á. (2025). El desarrollo del pensamiento espacial: una estrategia didáctica con Enfoque STEAM. *Fedumar Pedagogía y Educación, 12*(1), 90-97. https://doi.org/10.31948/fpe.v12i1.4938

Schoemann, A., Boulton, A., & Short, S. (2017). Determining Power and Sample Size for Simple and Complex Mediation Models. *Social Psychological and Personality Science, 8*(4), 379-386. https://doi.org/10.1177/1948550617715068

Tserendolgor, T., & Choijilsuren, N. (2026). A Method for Localizing the Spatial Thinking Ability Test (STAT): Adaptation and Validation of STAT-24 in Mongolia. *Journal of Geography*, *125*(5), 230-237. https://doi.org/10.1080/00221341.2026.2635037

Ulbricht, J., & Lindau, A. (2025). Pre-Service Geography Teachers' Spatial Conceptions—How Do Field Trips Shape Spatial Conceptions? *KN - Journal of Cartography and Geographic Information, 75*(4), 245-254. https://doi.org/10.1007/s42489-025-00208-x

Uttal, D., Meadow, N., Tipton, E., Hand, L., Alden, A., Warren, C., & Newcombe, N. (2013). The malleability of spatial skills: A meta-

analysis of training studies. *Psychological Bulletin, 139*(2), 352-402. https://doi.org/10.1037/a0028446

Von Reumont, F., & Budke, A. (2021). Spatial Thinking With Comics in Geography Education. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.702738

Wei, H., & Zhao, Y. (2026). A Study on a Geography Project-Based Learning Model Aimed at Cultivating Spatial Thinking. *Journal of Geography*, 1-16. https://doi.org/10.1080/00221341.2026.2667278

Woodley, E., Barr, S., Batty, L., Bickerstaff, K., Darvill, C., Ferdous, R., Holmes, N., Jon, I., Lynch, K., Martin, J., Marvell, A., McDougall, D., Pitt, H., Singleton, A., Souch, C., & Yorke, L. (2024). The future of geography field course pedagogy in UK higher education. *Geo: Geography and Environment, 11*(2). https://doi.org/10.1002/geo2.158

Xiang, X., & Xi, D. (2025). Examining cognitive processes of spatial thinking in university students: Insights from a web-based geographic information systems study. *British Journal of Educational Technology, 56*(1), 296-317. https://doi.org/10.1111/bjet.13502

Xie, S., Zheng, X., Sun, Y., Wan, J., & Lu, X. (2021). The Factors and Mechanisms That Influence Geospatial Thinking: A Structural Equation Modeling Approach. *Journal of Geography, 120*(5), 165-175. https://doi.org/10.1080/00221341.2021.1967423

CAPÍTULO XIII

PEDAGOGÍA DEL MICRORRELATO Y NARRATIVAS TRANSMEDIA EN LA FORMACIÓN DOCENTE EN LENGUA Y LITERATURA: UN MODELO PLS-SEM CON ANÁLISIS MULTIGRUPO EN EL PERÚ, 2025

Eugenio Ccanto Mallma

Universidad Nacional Intercultural de la Amazonía

Resumen

El microrrelato se emplea cada vez más en la enseñanza de la literatura, pero se desconoce mediante qué mecanismo su diseño pedagógico se traduce en competencia didáctica en quienes se forman como docentes. Este estudio contrasta un modelo de ecuaciones estructurales por mínimos cuadrados parciales en el que el diseño pedagógico del microrrelato incide sobre la competencia didáctica literaria de forma directa y a través de la competencia narrativa transmedia, y examina su heterogeneidad entre universidades y escuelas de educación superior pedagógica del Perú mediante análisis multigrupo condicionado a la invarianza de medición. Se aplicó un diseño transversal explicativo-predictivo a 346 docentes en formación (179 universitarios y 167 de escuelas pedagógicas), con tamaño muestral calculado a priori. El modelo de medida cumplió los criterios de fiabilidad y validez. El diseño pedagógico se asoció con la competencia didáctica literaria ($\beta = .333$) y con la competencia transmedia ($\beta = .668$); la mediación fue parcial (efecto indirecto = .176; VAF = .346) y el modelo mostró capacidad predictiva fuera de muestra. La relación con la competencia transmedia fue más intensa en las universidades (diferencia = −.127; $p = .030$), mientras que la relación con la competencia didáctica no difirió, aunque esta última carece de invarianza composicional. Los resultados sugieren que la competencia transmedia es una vía parcial de transferencia, y que la heterogeneidad institucional se concentra en la primera etapa del mecanismo.

Palabras clave: microrrelato; narrativa transmedia; formación docente; educación literaria; PLS-SEM; análisis multigrupo

Abstract

Flash fiction is increasingly used in literature teaching, yet the mechanism through which its pedagogical design becomes didactic competence among prospective teachers remains unknown. This study tests a partial least squares structural equation model in which the pedagogical design of flash fiction affects literary

didactic competence directly and through transmedia narrative competence, and it examines heterogeneity between universities and higher-education pedagogical schools in Peru through multigroup analysis conditional on measurement invariance. A cross-sectional, explanatory-predictive design was applied to 346 prospective teachers (179 university students and 167 from pedagogical schools), with sample size determined a priori. The measurement model met reliability and validity criteria. Pedagogical design was associated with literary didactic competence (β = .333) and with transmedia competence (β = .668); mediation was partial (indirect effect = .176; VAF = .346), and the model showed out-of-sample predictive capacity. The relationship with transmedia competence was stronger in universities (difference = −.127; p = .030), whereas the relationship with didactic competence did not differ, although the latter lacks compositional invariance. The findings suggest that transmedia competence is a partial transfer pathway and that institutional heterogeneity is concentrated in the first stage of the mechanism.

Keywords: *flash fiction; transmedia storytelling; teacher education; literary education; PLS-SEM; multigroup analysis*

13.1 INTRODUCCIÓN

La enseñanza de la literatura atraviesa una reconfiguración profunda. Las prácticas de lectura y escritura de las nuevas generaciones se desarrollan en ecologías mediáticas donde los relatos circulan, se fragmentan y se expanden a través de plataformas, lenguajes y comunidades de lectores, mientras la formación inicial del profesorado sigue organizada, en buena medida, en torno a la lectura analítica del texto impreso (Selfa y Falguera, 2022; Meyerhofer et al., 2024). Las revisiones sistemáticas recientes coinciden en que la narrativa transmedia ya no es un fenómeno periférico de la cultura popular, sino un campo educativo con objetos, prácticas y competencias propias, aunque su incorporación a la formación docente permanece fragmentaria y escasamente evaluada con evidencia cuantitativa (Meyerhofer, 2025; Meyerhofer y González, 2025). La revisión de la literatura sobre aprendizaje transmedia ha identificado cinco áreas clave —el juego transmedia, la cultura participativa, las habilidades transmedia,

la conexión y el aprendizaje informal— que enlazan el aprendizaje formal con el informal entre plataformas (Tombleson, 2024). En América Latina, el debate conceptual sobre la alfabetización transmedia y el estado de la cuestión sobre alfabetizaciones múltiples e interactividad muestran un campo en consolidación, todavía atravesado por tensiones teóricas (Pérez, 2023; Farfán, 2026).

En ese escenario, el microrrelato ocupa una posición singular. Su brevedad extrema, la condensación, la elipsis y el final abierto lo configuran como forma literaria distinta, que exige al lector una cooperación interpretativa intensa (Roche, 2024) y que, por su escala, se aviene con la lógica de la cultura de la convergencia, donde los relatos se distribuyen entre medios y cada uno aporta al universo narrativo (Jenkins, 2006). El género ha ganado reconocimiento académico y presencia en los manuales escolares de Lengua y Literatura, donde su definición y su canon constituyen todavía un asunto de debate curricular (Bartolomé, 2023), y se ha empleado incluso para articular la reflexión gramatical con la comprensión textual (Sánchez, 2025). Las experiencias didácticas documentadas muestran su potencial en entornos digitales, como los microrrelatos hipermediales concebidos para la enseñanza de la fantasía (Mateos y Álvarez, 2021).

Paralelamente, la creación multimodal y transmedia se ha incorporado a la escuela y a la educación superior. Se han descrito experiencias de relato transmedia lúdico que amplían las alfabetizaciones del alumnado (Colvert, 2022), estudios piloto sobre innovación y aprendizaje con productos transmedia en Colombia (García et al., 2024), enfoques transmedia para la educación literaria en los grados de formación de maestros (Clavijo, 2026), experiencias que integran literatura y transmedia en el nivel medio superior en México (García, 2025), estrategias didácticas multidisciplinarias articuladas en torno a una narrativa transmedia (Fernández et al., 2025) y prácticas de producción de vídeo como educación literaria multimodal (Allué y Cassany, 2023). Estas aproximaciones se apoyan en una concepción de la escritura como diseño, en la que el significado resulta de la orquestación de modos semióticos (Moses y Serafini, 2022), y en la idea de que los lectores noveles necesitan aprender a interpretar ficciones

transmedia (Mackey, 2025). Sin embargo, la literatura disponible describe experiencias de aula o propuestas de diseño; no modela empíricamente cómo el diseño pedagógico del microrrelato se transfiere a la competencia didáctica literaria de quienes se forman como docentes, ni por qué mecanismo lo hace.

El Perú ofrece un sitio de prueba especialmente informativo para esta pregunta. La formación inicial docente se imparte en dos tipos de institución —universidades y escuelas de educación superior pedagógica— con tradiciones curriculares distintas, y las políticas que la regulan han sido analizadas críticamente en clave comparada con las de Bolivia (Chuquilin, 2026). Además, el país cuenta con una tradición propia de microrrelato y minificción, visible tanto en la producción antológica reciente (Llerena, 2023) como en la obra de autores canónicos de la minificción peruana (Doria, 2023). Esta dualidad institucional permite examinar si las relaciones entre diseño pedagógico y competencias docentes son estructuralmente equivalentes entre contextos formativos o dependen de la institución que forma.

El presente estudio propone y contrasta un modelo de ecuaciones estructurales por mínimos cuadrados parciales en el que el diseño pedagógico del microrrelato incide sobre la competencia didáctica literaria directamente y a través de la competencia narrativa transmedia del docente en formación, y examina la heterogeneidad de esas relaciones entre tipos de institución mediante análisis multigrupo con contraste previo de invarianza de medición. Se formulan cuatro hipótesis:

1) H1, el diseño pedagógico del microrrelato se asocia positivamente con la competencia didáctica literaria;

2) H2, el diseño pedagógico del microrrelato se asocia positivamente con la competencia narrativa transmedia;

3) H3, la competencia narrativa transmedia media parcialmente la relación entre el diseño pedagógico del microrrelato y la competencia didáctica literaria; y

4) H4, la relación entre el diseño pedagógico del microrrelato y la competencia didáctica literaria difiere entre universidades y escuelas de educación superior pedagógica.

La contribución es de alcance general. En el plano teórico, el estudio especifica un mecanismo —la competencia narrativa transmedia como vía de transferencia— que conecta la didáctica de los géneros breves con la teoría de la multimodalidad, y que puede contrastarse en cualquier sistema de formación docente con géneros literarios breves y entornos digitales. En el plano metodológico, aporta un procedimiento replicable que combina el cálculo a priori del tamaño muestral para PLS-SEM, la evaluación de la invarianza composicional y la comparación multigrupo por permutación, de modo que la heterogeneidad entre instituciones se interprete solo donde la medición es comparable.

13.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

La teoría de la multimodalidad entiende el significado como resultado de la orquestación de modos semióticos —texto, imagen, sonido, disposición espacial— y concibe la escritura como diseño, no solo como codificación lingüística (Moses y Serafini, 2022). Desde esta perspectiva, enseñar microrrelato implica diseñar experiencias en las que la condensación verbal dialoga con otros modos y en las que el relato breve funciona como núcleo expandible. La cultura de la convergencia añade que los relatos contemporáneos se distribuyen entre medios, cada uno de los cuales aporta una contribución específica al universo narrativo (Jenkins, 2006), y la investigación educativa reciente ha trasladado esa lógica a la noción de alfabetización transmedia como conjunto de competencias de producción, interpretación y participación (Meyerhofer, 2025; Mackey, 2025).

1) El primer argumento sostiene que un diseño pedagógico deliberado del microrrelato —secuencias de lectura y escritura, combinación de modos, trabajo explícito con la poética de la brevedad y evaluación con criterios literarios— constituye una práctica auténtica de enseñanza de la literatura en la que el docente en formación aprende a enseñar

haciendo. Esa práctica debería traducirse en una mayor competencia didáctica literaria, entendida como la capacidad de seleccionar textos pertinentes, promover interpretaciones fundamentadas, articular lectura y creación, y evaluar coherentemente (Selfa y Falguera, 2022). De ahí la H1: el diseño pedagógico del microrrelato se asocia positivamente con la competencia didáctica literaria.

2) El segundo argumento sostiene que el mismo diseño, al exigir que el relato se proyecte en formatos visuales, audiovisuales y participativos, desarrolla la competencia narrativa transmedia del docente en formación (Colvert, 2022; García et al., 2024), lo que fundamenta la H2. La competencia narrativa transmedia, a su vez, funciona como recurso profesional: quien domina la expansión de un relato entre plataformas dispone de más repertorios para diseñar tareas literarias y para vincular la lectura con la producción del estudiantado. La evidencia sobre creatividad docente muestra que los efectos de las prácticas de enseñanza sobre la escritura creativa se transmiten a través de capacidades y creencias del propio docente (Hu y Choi, 2023; Hodges et al., 2021), lo que sustenta una mediación parcial: el diseño pedagógico actúa directamente y también a través de la competencia narrativa transmedia (H3).

3) El tercer argumento es institucional. Las universidades y las escuelas de educación superior pedagógica difieren en la organización del currículo, en la carga relativa de la formación disciplinar y de la práctica, y en la exposición a proyectos de innovación tecnológica. La investigación con modelos estructurales en formación inicial docente ha documentado que las relaciones entre creencias pedagógicas y adopción de innovaciones no son homogéneas entre poblaciones de futuros docentes (Gurer y Akkaya, 2022). En consecuencia, cabe esperar que la relación entre diseño pedagógico y competencia didáctica literaria difiera entre tipos de institución (H4). Esta comparación solo es interpretable si los constructos significan lo mismo en ambos grupos, razón por la cual el análisis multigrupo se condiciona a la invarianza de medición de los compuestos (Henseler et al., 2016).

13.3 MATERIALES Y MÉTODOS

13.3.1 Diseño y participantes

El estudio adoptó un diseño cuantitativo, no experimental, transversal y explicativo-predictivo, con recolección de datos durante el segundo semestre académico de 2025. La unidad de análisis fue el docente en formación inicial de la especialidad de Comunicación o de Lengua y Literatura de universidades y de escuelas de educación superior pedagógica del Perú.

El tamaño muestral se determinó a priori con el método de la raíz cuadrada inversa para PLS-SEM (Kock y Hadaya, 2018), con $\alpha = .05$ unilateral y potencia de .80, de modo que se detectara un coeficiente de ruta de .15 en la muestra total y uno de .20 en cada grupo del análisis multigrupo, valor este último que corresponde al umbral de relevancia propuesto para modelos PLS (Chin, 1998).

Con una proporción esperada de .50 por institución, N = 346 es el menor tamaño total para el que la probabilidad de que el grupo menor quede por debajo del mínimo es inferior a .05 (Tabla 13-1). El método no depende del número de predictores; el constructo con más rutas entrantes, la competencia didáctica literaria, tiene dos, de modo que el mínimo de grados de libertad queda cubierto en cada grupo. La composición alcanzada fue de 179 docentes en formación de universidades y 167 de escuelas de educación superior pedagógica.

Tabla 13-1

Cálculo a priori del tamaño muestral (método de la raíz cuadrada inversa)

Criterio	Parámetros	Cálculo	Resultado
Muestra total	$\beta_min = .15$; $\alpha = .05$ unilateral; $1 - \beta = .80$	$n = (2.4865 / 0.15)^2 = 274.8$	275
Cada grupo del análisis multigrupo	$\beta_min = .20$; $\alpha = .05$ unilateral; $1 - \beta = .80$	$n = (2.4865 / 0.20)^2 = 154.6$	155 por grupo
Composición por institución	proporción esperada .50; n por grupo ≥ 155	$P = 2 \cdot F_Bin(154; N, .50) \leq .05 \rightarrow N = 346$ ($P = .0465$); $N = 345$ ($P = .0524$)	N = 346
Composición alcanzada	Universidad / Escuela de educación superior pedagógica	179 ≥ 155; 167 ≥ 155	Requisito cumplido en ambos grupos

Nota: $2.4865 = z_{0.95} + z_{0.80} = 1.6449 + 0.8416$. F_Bin = función de distribución binomial acumulada. El tamaño adoptado cubre el análisis más exigente (comparación multigrupo).

La selección fue no probabilística, por conveniencia, entre instituciones que facilitaron el acceso a estudiantes de los últimos ciclos de la especialidad; por ello, los resultados describen las relaciones observadas en los participantes analizados y no se extrapolan a la población nacional. Se excluyeron los cuestionarios incompletos y los patrones de respuesta invariantes.

13.3.2 Instrumentos

Los tres constructos se midieron como compuestos reflectivos (modo A) con escalas de autoinforme de cuatro ítems y respuesta tipo Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo), formato afín al de los instrumentos validados con docentes y estudiantes de pedagogía (Isshiki, 2026; Nabhan y Habók, 2025; Hodges et al., 2021). La operacionalización de la competencia narrativa transmedia se apoyó en la medición de la alfabetización transmedia en estudiantes universitarios (Rivera, 2024) y la del diseño pedagógico, en experiencias de formación docente en competencias digitales para Lengua y Literatura (Jordan, 2023). El tipo de institución formadora se registró como variable de agrupación (Tabla 13-2).

Tabla 13-2

Operacionalización de los constructos y de la variable de agrupación

Constructo	Código	Ítem	Tipo de medida
Diseño pedagógico del microrrelato	DPM1	Diseño secuencias didácticas de lectura y escritura basadas en microrrelatos	Reflectivo, Likert 1–5
	DPM2	Combino texto, imagen y sonido al planificar actividades con microrrelatos	
	DPM3	Uso el microrrelato para trabajar la condensación, la elipsis y el final abierto	
	DPM4	Evalúo la producción de microrrelatos con criterios literarios explícitos	
Competencia narrativa transmedia	CNT1	Expando un relato breve a otras plataformas y lenguajes sin perder su coherencia	
	CNT2	Diseño recorridos narrativos que integran redes sociales, audio y video	
	CNT3	Adapto un microrrelato a formatos visuales y audiovisuales	
	CNT4	Planifico la participación de los lectores en la expansión de un relato	
Competencia didáctica literaria	CDL1	Selecciono textos literarios pertinentes para el nivel y el contexto de mis estudiantes	
	CDL2	Promuevo la interpretación fundamentada de textos literarios breves	
	CDL3	Articulo la lectura literaria con la producción creativa de los estudiantes	
	CDL4	Evalúo la competencia literaria con instrumentos coherentes con los objetivos	
Institución formadora	GRP	Universidad (n = 179) / Escuela de educación superior pedagógica (n = 167)	Agrupación dicotómica

Nota: DPM = diseño pedagógico del microrrelato; CNT = competencia narrativa transmedia; CDL = competencia didáctica literaria. Los ítems se estandarizaron (puntuaciones z) antes de la estimación.

13.3.3 Análisis de datos

Se estimó un modelo PLS-SEM con compuestos en modo A, por su orientación explicativo-predictiva (Hair et al., 2021). La significación se evaluó con bootstrap de 5000 submuestras e intervalos percentiles al 95 %.

El modelo de medida se valoró con cargas externas ($\lambda \geq .708$), fiabilidad (α de Cronbach, ρA y ρc entre .70 y .95) y AVE ($\geq .50$); la validez discriminante, con el criterio de Fornell-Larcker, el HTMT (< .85; Henseler et al., 2015) y las cargas cruzadas. El modelo estructural se valoró con coeficientes de ruta, VIF interno, R^2, f^2, SRMR (< .08) y PLSpredict con validación cruzada de 10 particiones frente a un modelo lineal de referencia (Hair et al., 2021; Sharma et al., 2023). El efecto indirecto se estimó como a × b con bootstrap de 5000 réplicas, junto con el efecto directo c′, el efecto total y la proporción mediada VAF = a×b / (c′ + a×b). El sesgo de método común se examinó con el VIF de colinealidad completa (umbral 3.3; Kock, 2015).

La heterogeneidad se evaluó en dos pasos. Primero se contrastó la invarianza de medición de los compuestos con MICOM: invarianza configural por diseño, invarianza composicional mediante la correlación c entre puntuaciones calculadas con los pesos de cada grupo, e igualdad de medias y varianzas, con 5000 permutaciones (Henseler et al., 2016). Segundo, se estimó el modelo por grupo y se contrastaron las diferencias de los coeficientes de ruta y del efecto indirecto con la prueba de permutación PLS-MGA de 5000 permutaciones (Klesel et al., 2022). Como contraste complementario de H4 se estimó un término de interacción grupo × diseño pedagógico con errores estándar robustos HC3.

13.3 RESULTADOS

13.3.1 Modelo de medida y validez discriminante

Todas las cargas externas superaron el criterio de .708, y las fiabilidades y la varianza media extraída cumplieron los umbrales establecidos (Tabla 13-3).

Tabla 13-3

Modelo de medición: cargas externas, fiabilidad y validez convergente

Constructo	Ítem	λ	λ²	Peso externo	α	ρA	ρc	AVE
DPM	DPM1	.743	.552	.246	.819	.830	.880	.648
	DPM2	.816	.665	.329	.819	.830	.880	.648
	DPM3	.852	.726	.342	.819	.830	.880	.648
	DPM4	.806	.649	.319	.819	.830	.880	.648
CNT	CNT1	.825	.681	.263	.881	.887	.918	.737
	CNT2	.850	.722	.286	.881	.887	.918	.737
	CNT3	.869	.755	.290	.881	.887	.918	.737
	CNT4	.889	.790	.324	.881	.887	.918	.737
CDL	CDL1	.740	.547	.217	.843	.870	.894	.679
	CDL2	.804	.646	.283	.843	.870	.894	.679
	CDL3	.883	.779	.337	.843	.870	.894	.679
	CDL4	.861	.742	.366	.843	.870	.894	.679

Nota: λ = carga externa estandarizada del modelo PLS (modo A); λ² = fiabilidad del indicador; ρA = coeficiente de Dijkstra-Henseler; ρc = fiabilidad compuesta; AVE = varianza media extraída, calculados a partir de las cargas externas PLS. Criterios: λ ≥ .708; α, ρA y ρc entre .70 y .95; AVE ≥ .50.

La validez discriminante quedó establecida por los tres criterios: la raíz cuadrada del AVE de cada constructo superó sus correlaciones con los demás, todos los HTMT fueron inferiores a .85 y cada ítem cargó más en su constructo que en cualquier otro (Tabla 13-4; Figura 13-1).

Tabla 13-4

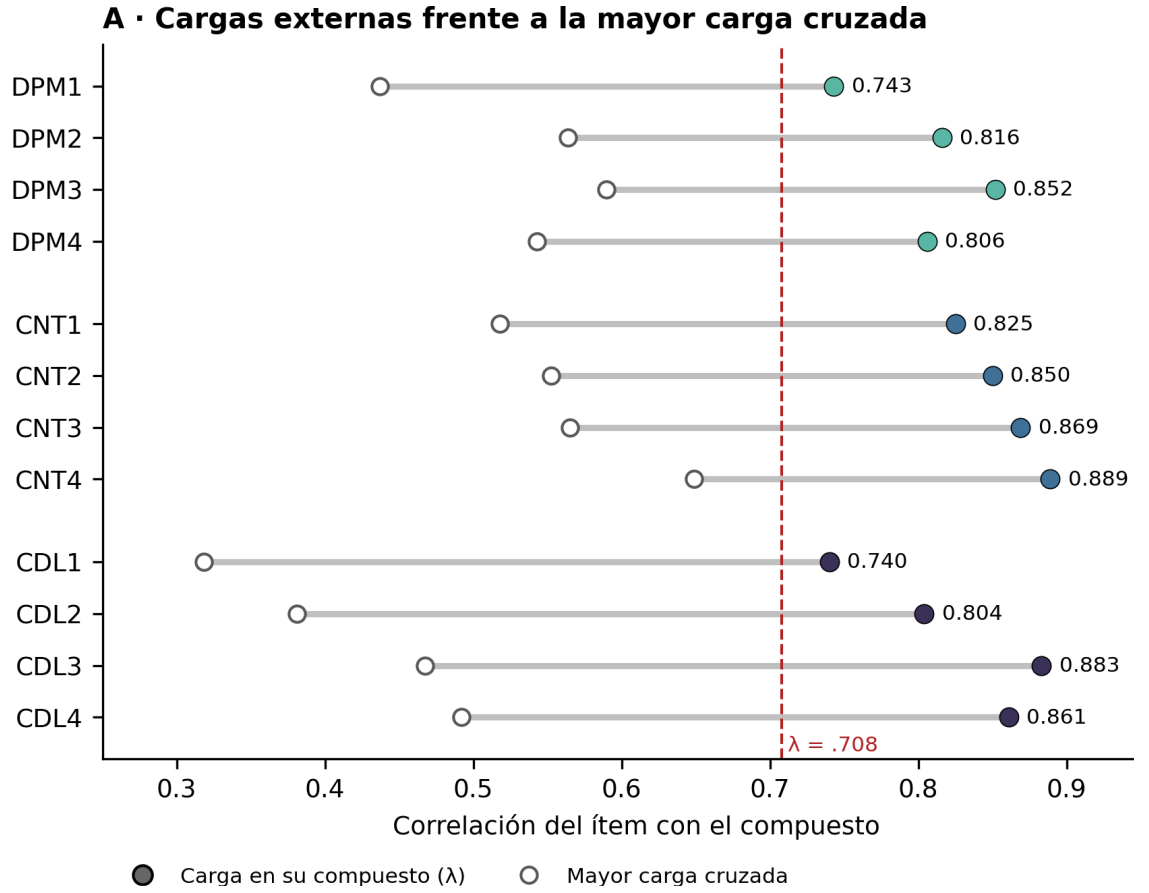
Estadísticos descriptivos, correlaciones entre constructos y validez discriminante

Variable	M	DE	Asimetría	Curtosis	DPM	CNT	CDL
DPM	0.017	0.790	−0.137	−0.336	**.805**	.776	.592
CNT	−0.009	0.868	−0.168	−0.004	.668*** [.605, .723]	**.859**	.545
CDL	0.014	0.832	0.004	0.733	.509*** [.426, .583]	.486*** [.401, .563]	**.824**

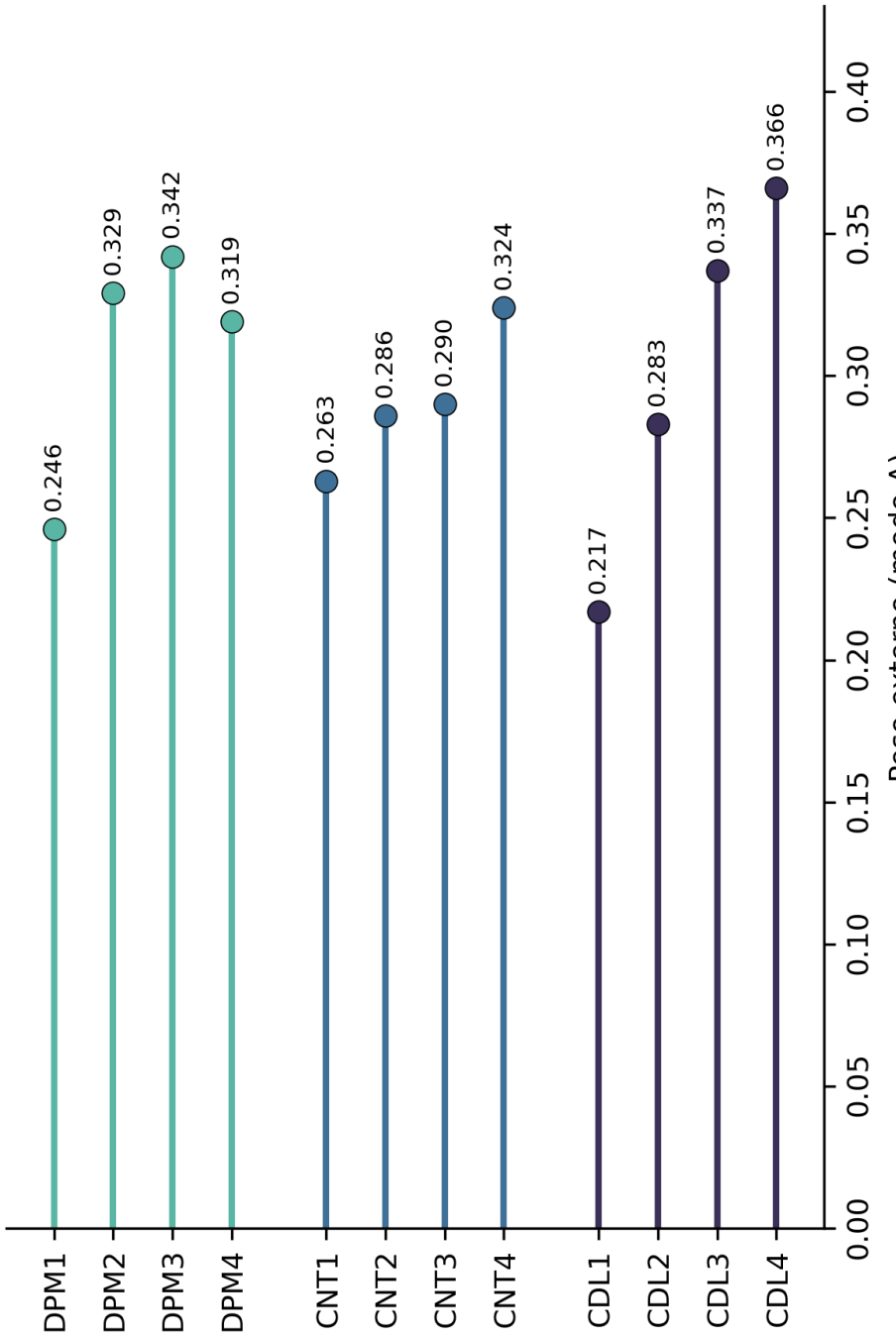
Nota: M y DE de la media de los ítems estandarizados de cada constructo. Diagonal en negrita: raíz cuadrada del AVE (criterio de Fornell-Larcker). Triángulo inferior: correlaciones entre las puntuaciones de los compuestos PLS con IC95 % (transformación Z de Fisher). Triángulo superior: HTMT. ***p < .001.

Figura 13-1

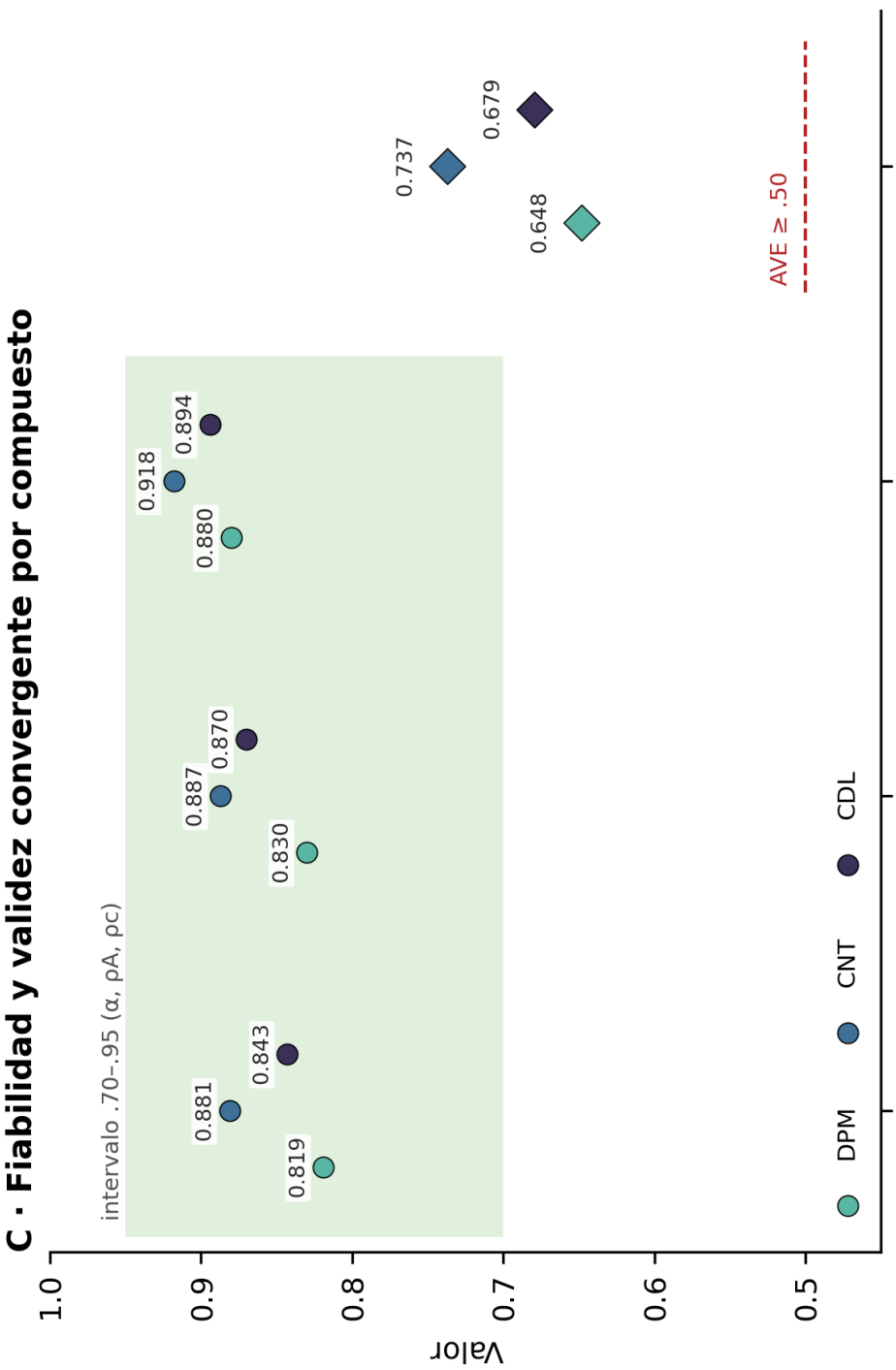
Arquitectura del modelo de medida de los compuestos



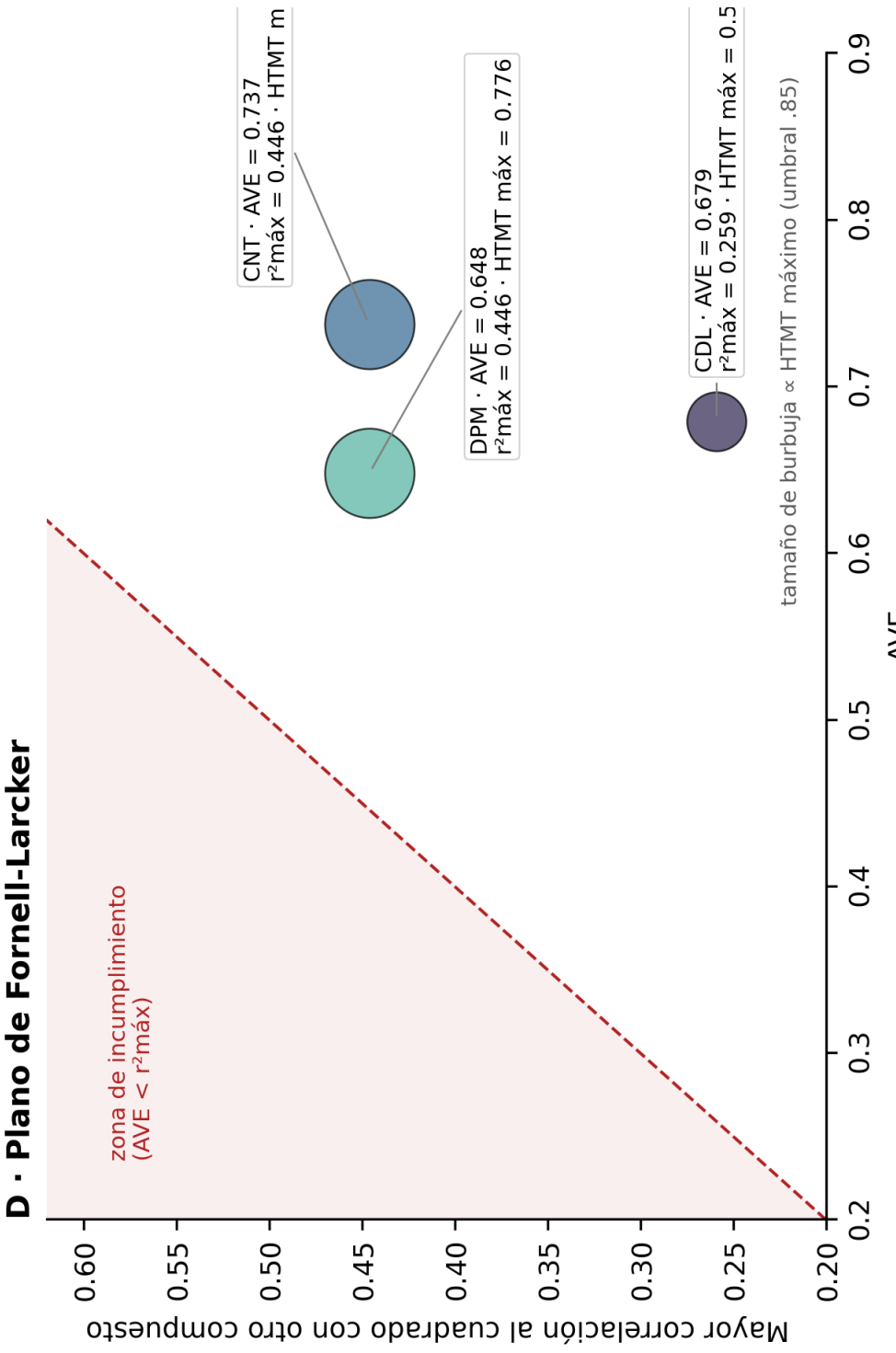
Nota. N = 346; PLS-SEM con compuestos en modo A; valores de las Tablas 3 y 4. Panel A: carga externa de cada ítem en su compuesto (rango .740–.889) frente a su mayor carga cruzada; la línea discontinua marca λ = .708.



Nota. N = 346; PLS-SEM con compuestos en modo A; valores de las Tablas 3 y 4. Panel B: pesos externos de los indicadores (rango .217–.366).



Nota. N = 346; PLS-SEM con compuestos en modo A; valores de las Tablas 3 y 4. Panel C: α de Cronbach (.819–.881), ρA (.830–.887), ρc (.880–.918) y AVE (.648–.737) por compuesto, con el intervalo .70–.95 sombreado y la referencia AVE ≥ .50.



Nota. N = 346; PLS-SEM con compuestos en modo A; valores de las Tablas 3 y 4. Panel D: AVE frente a la mayor correlación al cuadrado con otro compuesto (DPM .648 frente a .446; CNT .737 frente a .446; CDL .679 frente a .259); ningún compuesto cae en la zona de incumplimiento del criterio de Fornell-Larcker; el tamaño de la burbuja es proporcional al HTMT máximo (DPM y CNT .776; CDL .592; umbral .85). DPM = diseño pedagógico del microrrelato; CNT = competencia narrativa transmedia; CDL = competencia didáctica literaria.

13.3.2 Modelo estructural y mediación

El VIF interno máximo fue de 1.806. Las tres rutas del modelo resultaron positivas y significativas (Tabla 13-5), y el modelo explicó el 44.6 % de la varianza de CNT y el 29.7 % de la de CDL. La descomposición de efectos mostró un efecto indirecto significativo junto con un efecto directo significativo; la vía indirecta representó el 34.6 % del efecto total (Tabla 13-6; Figura 13-2).

Tabla 13-5

Coeficientes de ruta del modelo estructural PLS-SEM

Ruta	β	EE	t	p	IC 95 %	f^2	VIF
DPM → CNT (a)	.668	.029	23.108	< .001	[.609, .722]	.806	1.000
DPM → CDL (c′)	.333	.056	5.961	< .001	[.222, .441]	.087	1.806
CNT → CDL (b)	.264	.061	4.355	< .001	[.143, .383]	.055	1.806

Nota: β = coeficiente de ruta estandarizado; EE, t, p e IC95 % percentil por bootstrap de 5000 submuestras. f^2: .02 pequeño, .15 mediano, .35 grande. VIF = factor de inflación de la varianza interno.

Tabla 10-8

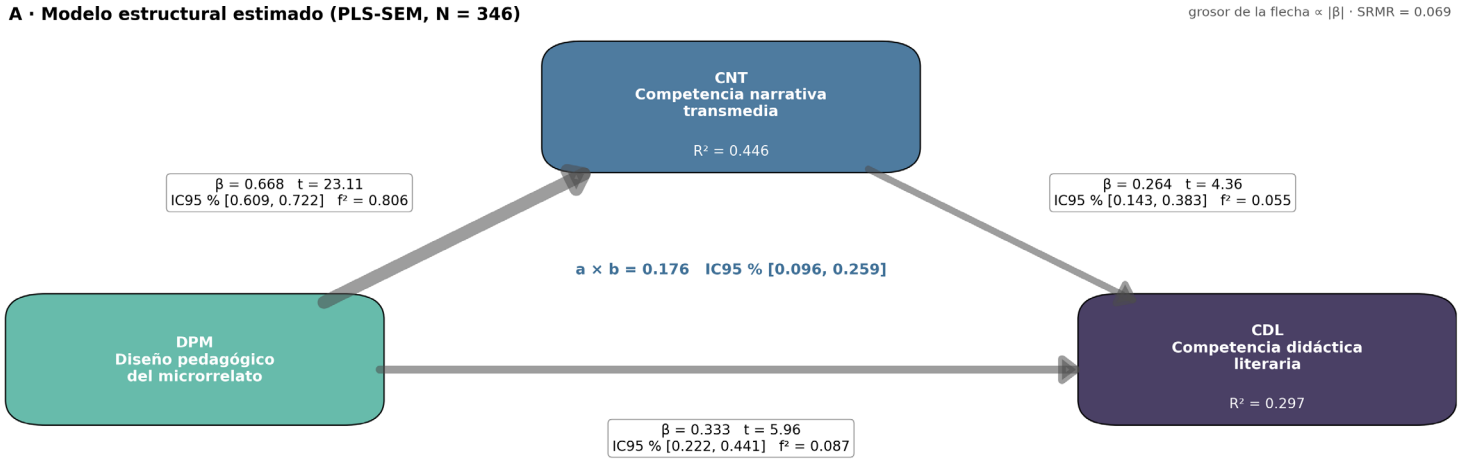
Descomposición de efectos: directo, indirecto y total

Efecto		Estimación	IC 95 %		Significativo
			inferior	superior	
a:	DPM → CNT	.668	.609	.722	Sí
b:	CNT → CDL	.264	.143	.383	Sí
c′:	DPM → CDL (directo)	.333	.222	.441	Sí
a × b:	DPM → CNT → CDL (indirecto)	.176	.096	.259	Sí
c:	efecto total (c′ + a × b)	.509	.436	.577	Sí

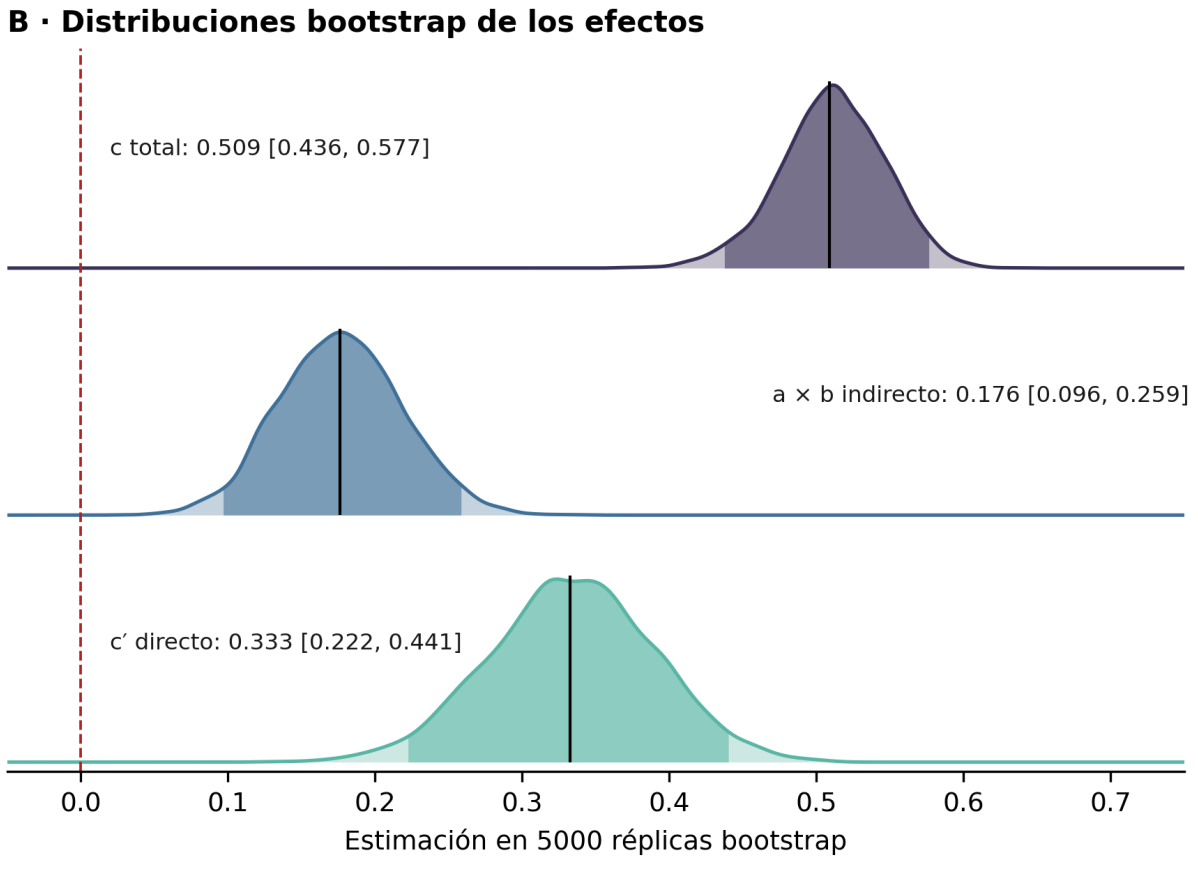
Nota: Intervalos percentiles por bootstrap de 5000 réplicas; un efecto es significativo cuando su intervalo excluye el cero. VAF = .176 / .509 = .346 (mediación parcial); proporción directa = .333 / .509 = .654 (65.4 %).

Figura 13-2

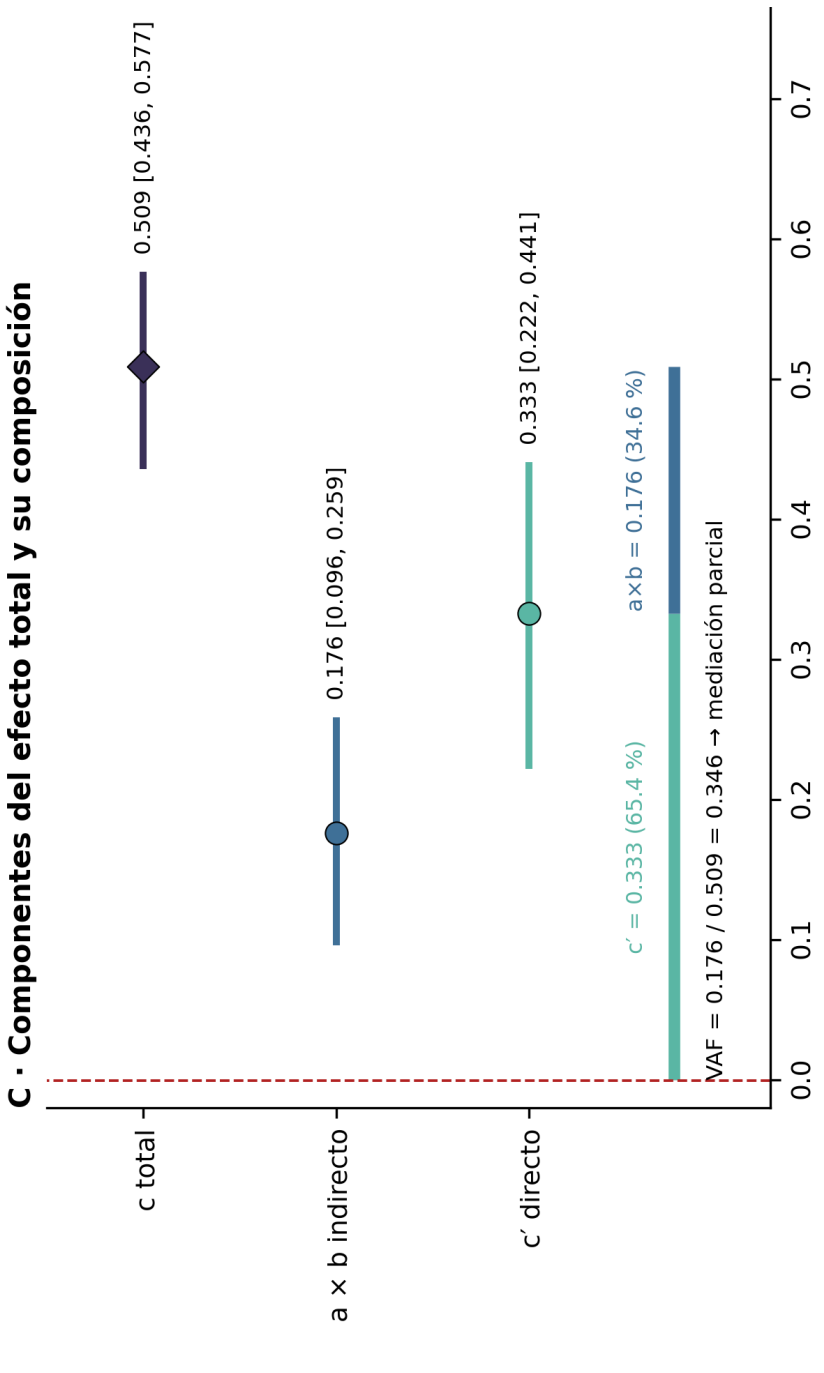
Modelo estructural estimado y descomposición del efecto del diseño pedagógico del microrrelato sobre la competencia didáctica literaria



Nota. N = 346; PLS-SEM con bootstrap de 5000 submuestras; valores de las Tablas 13-5 y 13-6. Panel A: coeficientes de ruta con t, IC95 % percentil y f^2: DPM → CNT β = .668 [.609, .722], f^2 = .806; DPM → CDL β = .333 [.222, .441], f^2 = .087; CNT → CDL β = .264 [.143, .383], f^2 = .055; R^2 = .446 (CNT) y .297 (CDL); efecto indirecto a × b = .176 [.096, .259]; SRMR = .069; el grosor de cada flecha es proporcional a |β|.



Nota. N = 346; PLS-SEM con bootstrap de 5000 submuestras; valores de las Tablas 13-5 y 13-6. Panel B: distribuciones bootstrap del efecto directo (c′ = .333 [.222, .441]), del indirecto (.176 [.096, .259]) y del total (.509 [.436, .577]), con el IC95 % percentil sombreado.



Nota. N = 346; PLS-SEM con bootstrap de 5000 submuestras; valores de las Tablas 13-5 y 13-6. Panel C: estimación puntual e IC95 % percentil del efecto directo, del indirecto y del total; VAF = a × b / c total = .176 / .509 = .346 (mediación parcial). DPM = diseño pedagógico del microrrelato; CNT = competencia narrativa transmedia; CDL = competencia didáctica literaria.

13.3.3 Capacidad explicativa y predictiva, y sesgo de método común

El ajuste aproximado fue adecuado, las predicciones fuera de muestra de CDL fueron relevantes y el error del modelo PLS resultó menor que el del modelo lineal de referencia (Tabla 13-7). Los VIF de colinealidad completa se situaron entre 1.506 y 2.496, por debajo del umbral de 3.3 (Tabla 13-8; Figura 13-3).

Tabla 13-7

Ajuste aproximado, capacidad explicativa y capacidad predictiva fuera de muestra

Indicador	Constructo	Valor	Referencia
SRMR	Modelo estimado	.069	< .08
R^2 (R^2 ajustado)	CNT	.446 (.445)	.25 débil; .50 moderado; .75 sustancial
R^2 (R^2 ajustado)	CDL	.297 (.293)	.25 débil; .50 moderado; .75 sustancial
Q^2predict	CDL	.274	> 0: relevancia predictiva
RMSE PLS / RMSE modelo lineal de referencia	CDL	.474 / .479	RMSE PLS < RMSE de referencia

Nota: PLSpredict con validación cruzada de 10 particiones; el modelo lineal de referencia regresa los indicadores endógenos sobre los indicadores exógenos.

Tabla 13-8

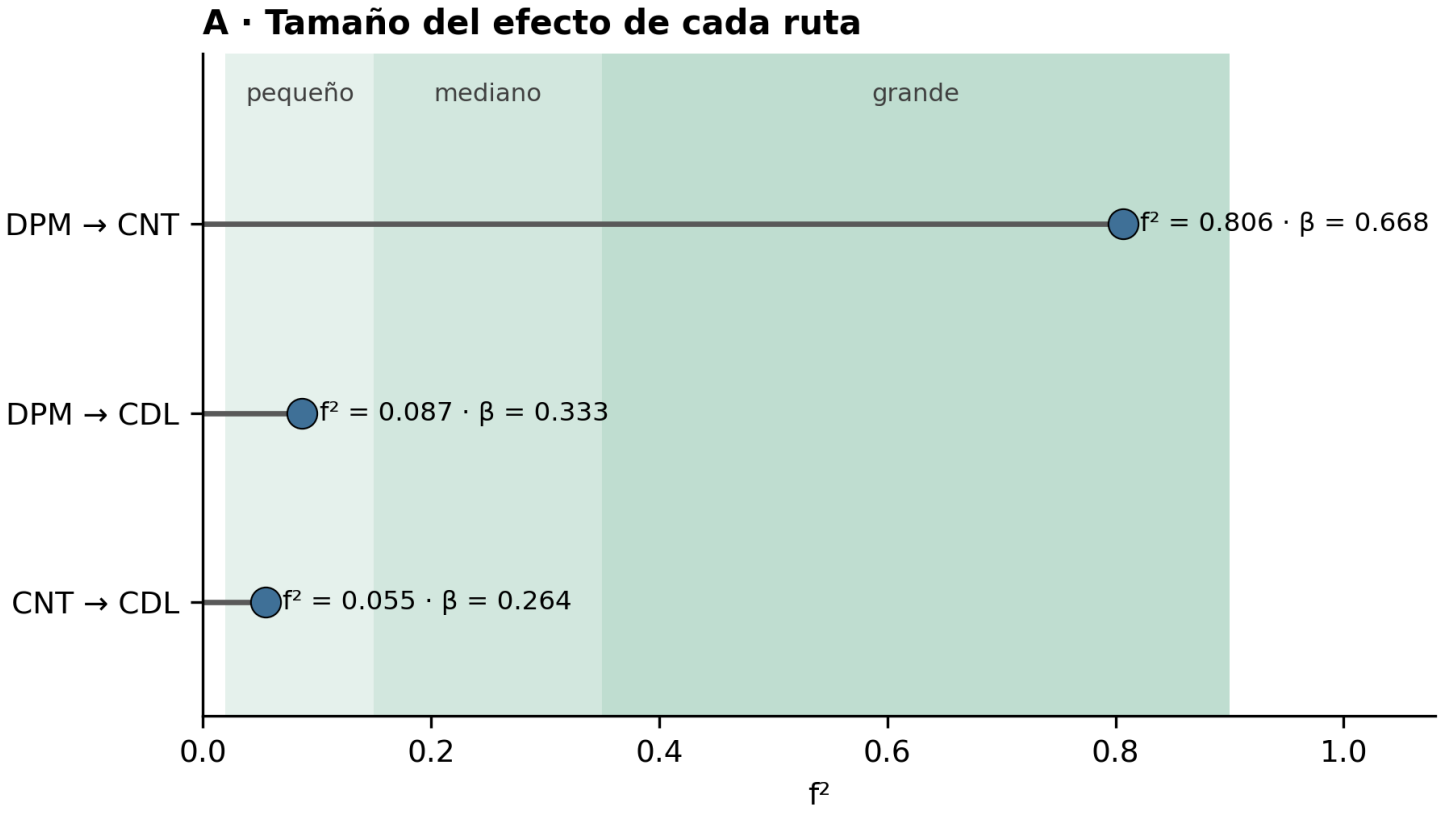
Sesgo de método común: factores de inflación de la varianza de colinealidad completa

Constructo	R^2 con los demás constructos	VIF de colinealidad completa	Criterio < 3.3
DPM	.585	2.410	Cumple
CNT	.599	2.496	Cumple
CDL	.336	1.506	Cumple

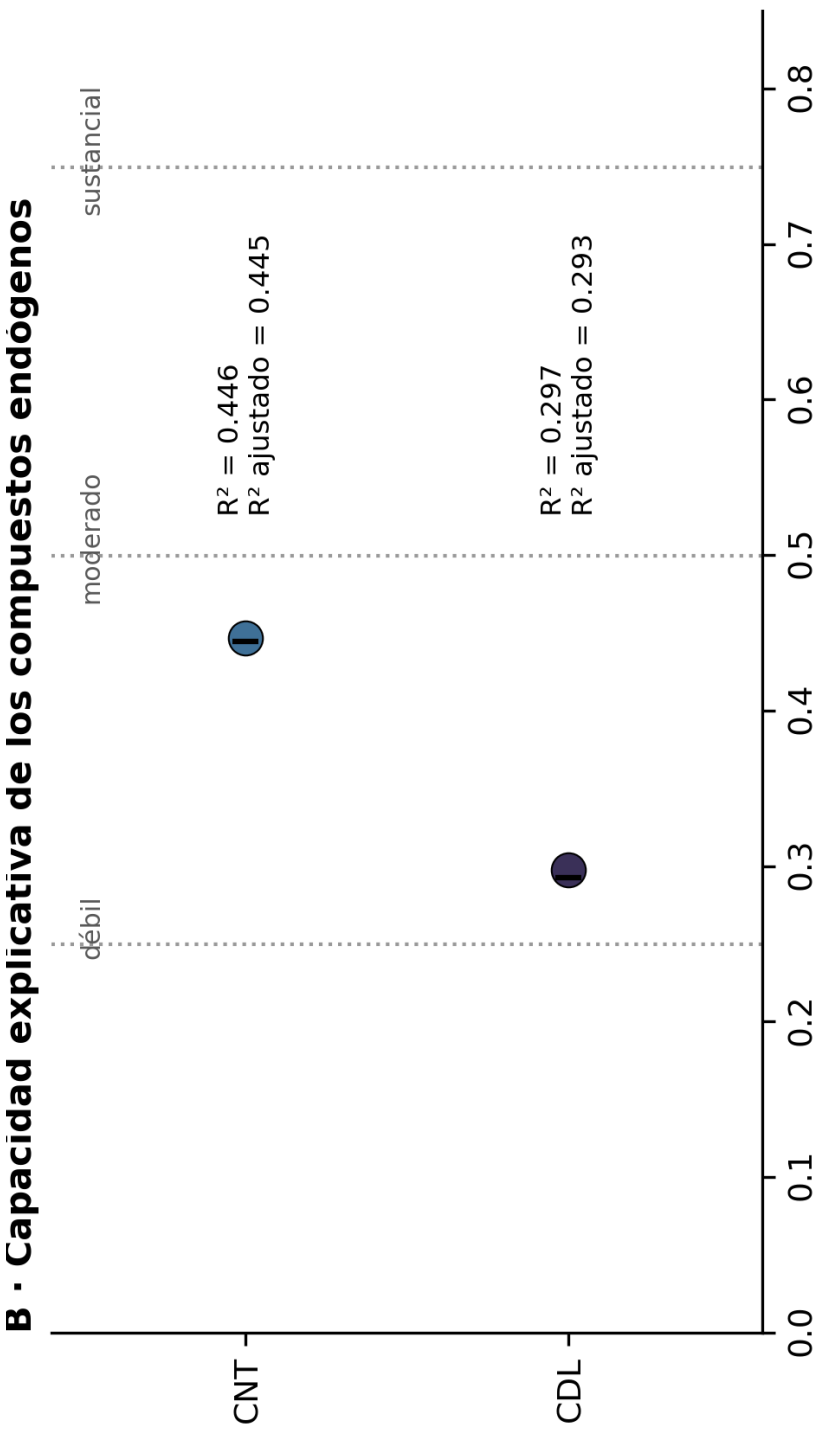
Nota: VIF = 1 / (1 – R^2), con cada compuesto regresado sobre los demás (Kock, 2015).

Figura 13-3

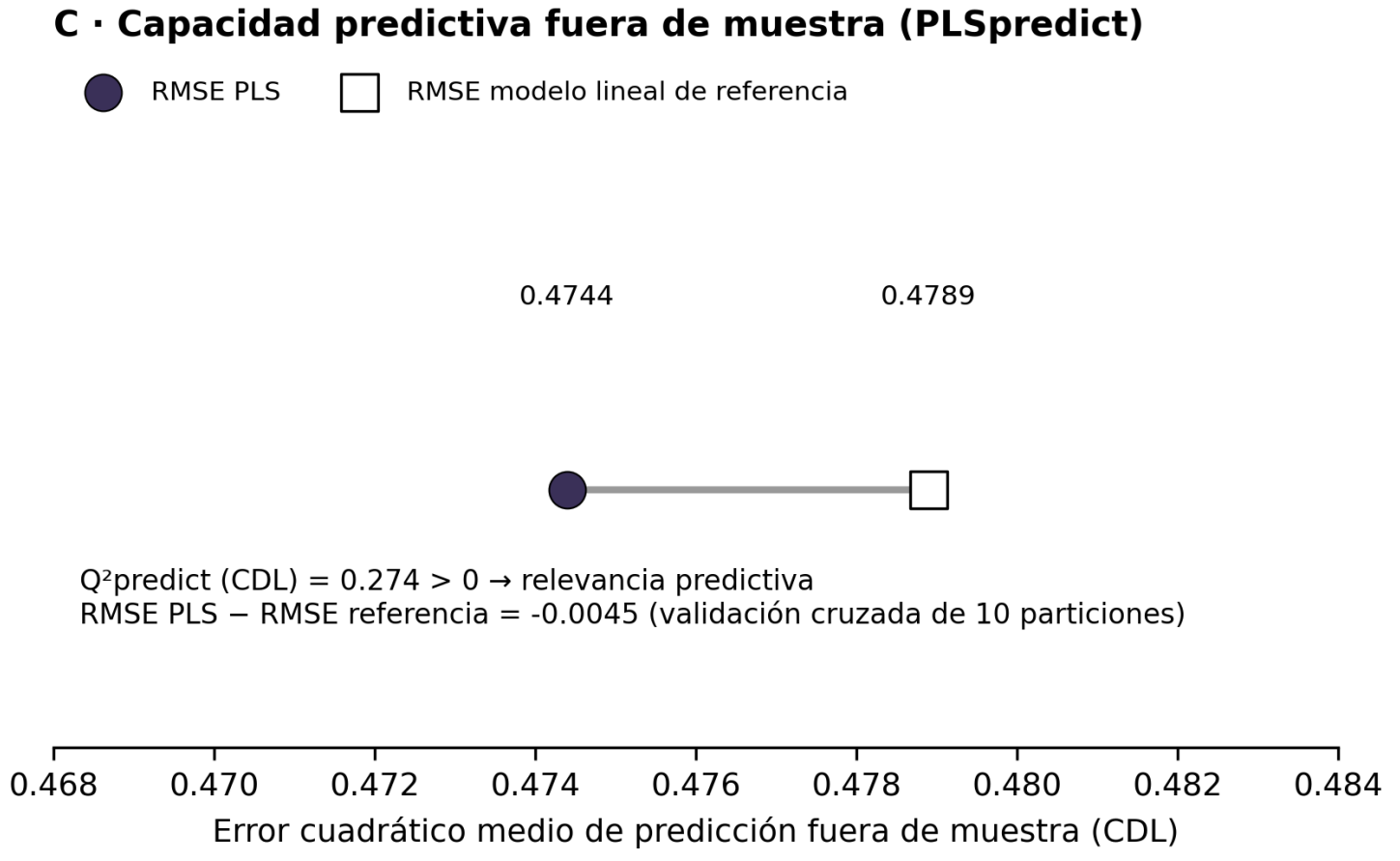
Tamaño del efecto, capacidad explicativa y predictiva, y sesgo de método común



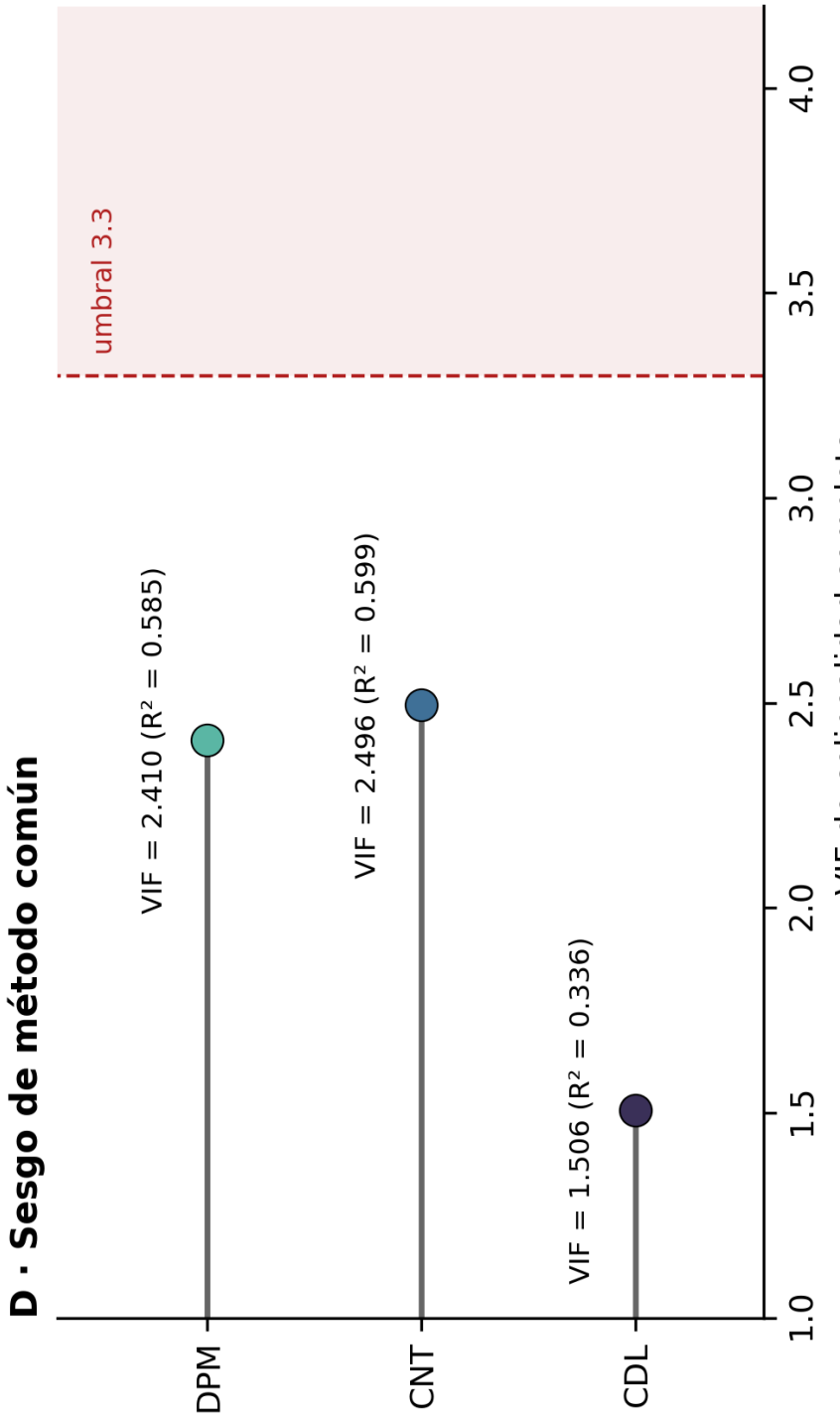
Nota. N = 346; valores de las Tablas 13-5, 13-7 y 13-8. Panel A: f² de cada ruta con las bandas convencionales (.02, .15 y .35): DPM → CNT .806; DPM → CDL .087; CNT → CDL .055.



Nota. N = 346; valores de las Tablas 13-5, 13-7 y 13-8. Panel B: R^2 y R^2 ajustado de CNT (.446 y .445) y de CDL (.297 y .293), con las referencias .25, .50 y .75.



Nota. N = 346; valores de las Tablas 13-5, 13-7 y 13-8. Panel C: error cuadrático medio de predicción fuera de muestra de CDL con PLSpredict y validación cruzada de 10 particiones: .4744 frente a .4789 del modelo lineal de referencia; Q^2predict = .274.



Nota. N = 346; valores de las Tablas 13-5, 13-7 y 13-8. Panel D: VIF de colinealidad completa de DPM (2.410), CNT (2.496) y CDL (1.506), todos por debajo de 3.3. DPM = diseño pedagógico del microrrelato; CNT = competencia narrativa transmedia; CDL = competencia didáctica literaria.

13.3.4 Invarianza de medición y análisis multigrupo

El paso composicional de MICOM estableció la invarianza en DPM y en CNT, con medias y varianzas iguales entre grupos, por lo que ambos compuestos presentan invarianza completa. En CDL, la correlación composicional quedó por debajo del percentil 5 de su distribución de permutación y la razón de varianzas se situó fuera de su intervalo, de modo que la comparación de las relaciones que desembocan en CDL no es interpretable entre grupos (Tabla 13-9).

Tabla 13-9

Invarianza de medición de los compuestos entre tipos de institución (MICOM)

Constructo	c	Percentil 5 de c	Invarianza composicional	Diferencia de medias [IC95 % de permutación]	Log razón de varianzas [IC95 % de permutación]	Resultado
DPM	.9987	.9975	Sí	0.052 [−0.219, 0.212]	0.011 [−0.275, 0.266]	Invarianza completa
CNT	.9999	.9991	Sí	0.068 [−0.222, 0.211]	−0.079 [−0.297, 0.286]	
CDL	.9950	.9963	No	0.184 [−0.217, 0.206]	0.384 [−0.357, 0.346]	Sin invarianza composicional

Nota: c = correlación entre las puntuaciones del compuesto calculadas con los pesos de cada grupo; la invarianza composicional se establece cuando c no es inferior al percentil 5 de su distribución bajo 5000 permutaciones. Diferencias calculadas como escuela de educación superior pedagógica menos universidad.

En la comparación multigrupo, la ruta DPM → CNT fue más intensa en las universidades que en las escuelas de educación superior pedagógica, con una diferencia significativa. La ruta DPM → CDL, la ruta b y el efecto indirecto no difirieron entre grupos (Tabla 13-10; Figura 13-4). El contraste complementario con un término de interacción grupo × diseño pedagógico y errores HC3 sobre los compuestos arrojó una diferencia de pendientes ($\Delta\beta = .137$, $p = .012$), que la prueba de permutación sobre el modelo PLS no confirmó.

Tabla 13-10

Análisis multigrupo PLS por permutación: universidades frente a escuelas de educación superior pedagógica

Ruta	Universidad (n = 179)	Escuela de educación superior pedagógica (n = 167)	Diferencia	p de permutación
	β [IC95 %]			
DPM → CNT (a)	.731 [.671, .788]	.604 [.493, .702]	−.127	.030
DPM → CDL (c′)	.277 [.106, .456]	.393 [.268, .530]	.116	.296
CNT → CDL (b)	.243 [.067, .409]	.302 [.144, .449]	.059	.621
DPM → CNT → CDL (indirecto)	.177 [.049, .304]	.182 [.087, .289]	.005	.957

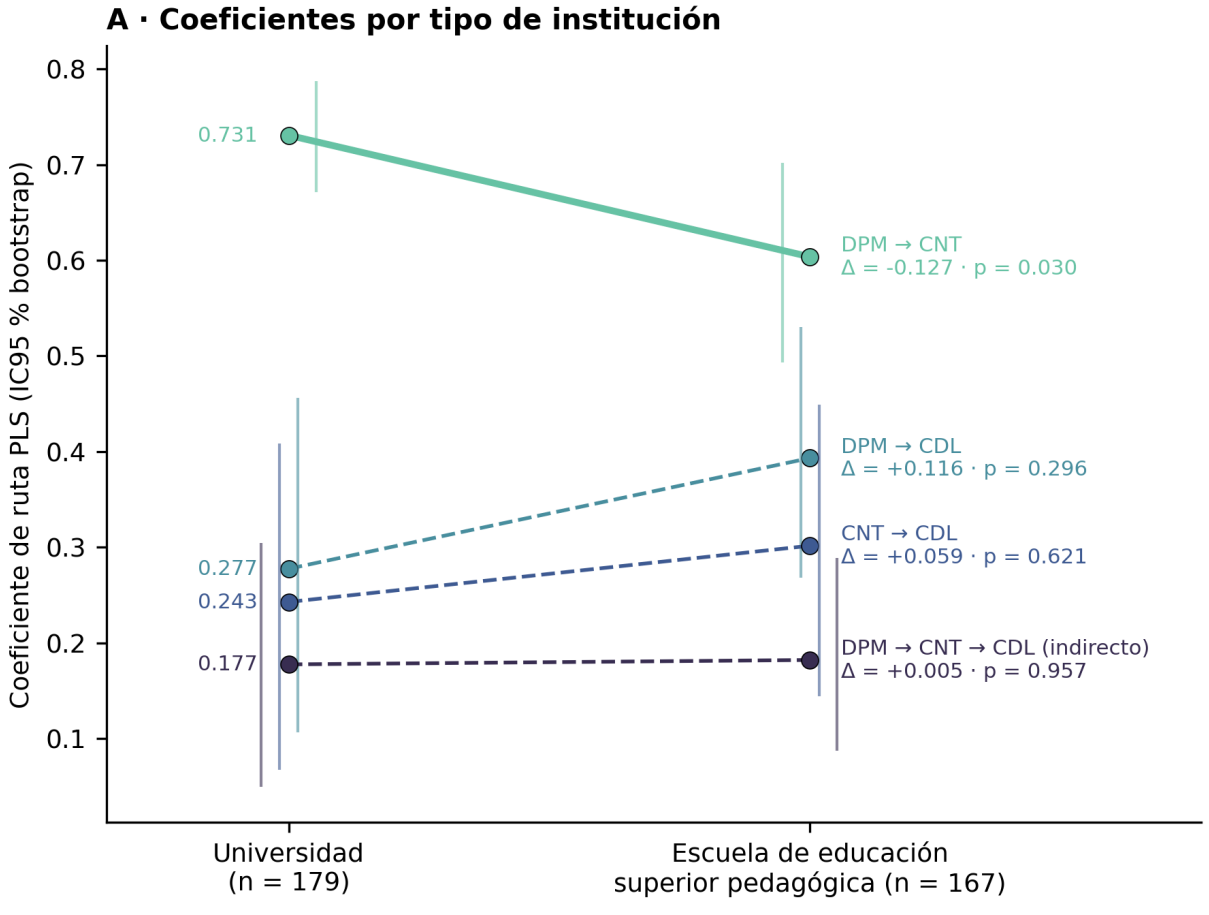
Nota: Coeficientes estimados por separado en cada grupo; IC95 % percentil por bootstrap de 5000 submuestras dentro de cada grupo. Diferencia = escuela de educación superior pedagógica menos universidad. p bilateral de la prueba de permutación PLS-MGA con 5000 permutaciones. Las comparaciones de las rutas que desembocan en CDL se reportan con la salvedad de la Tabla 13-9.

13.3.5 Soporte de las hipótesis

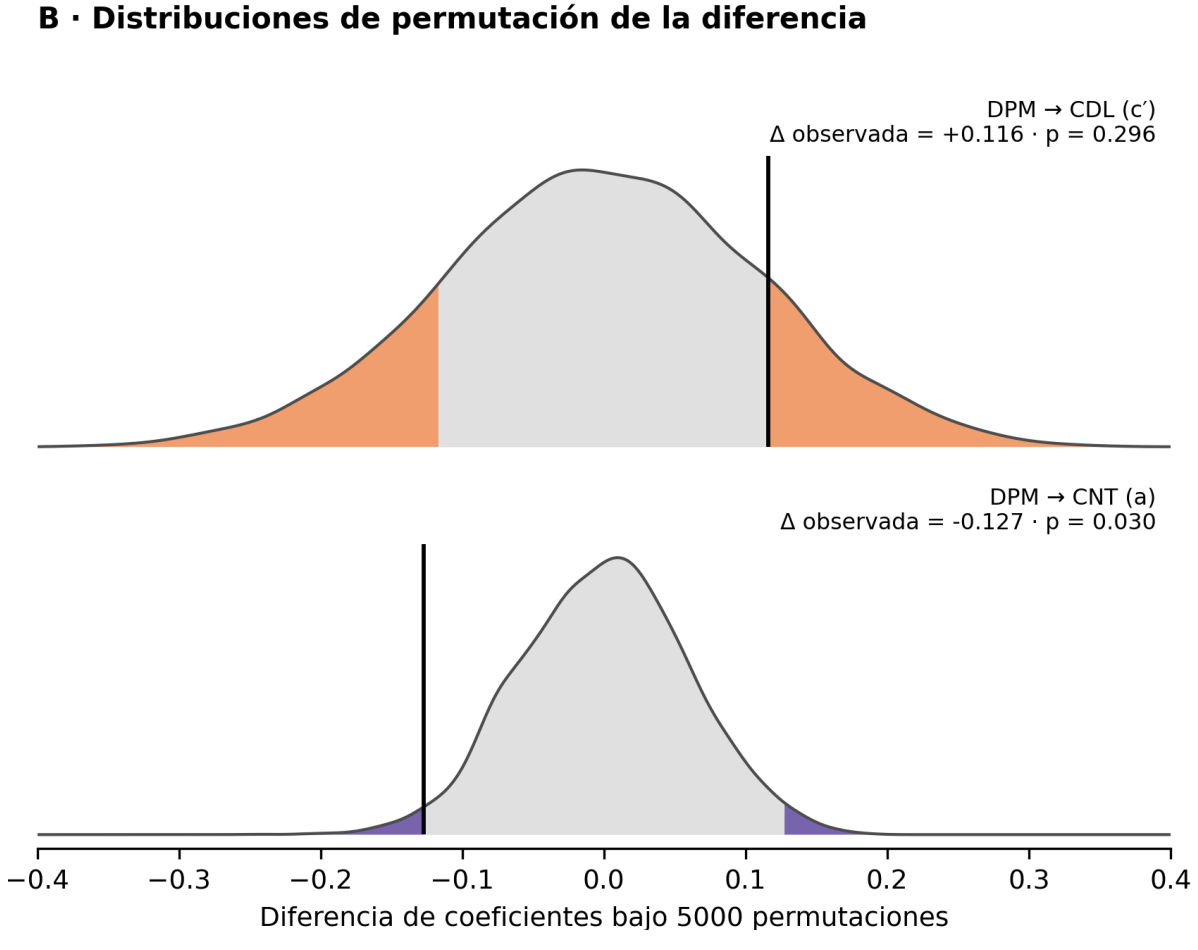
H1 (β = .333, p < .001), H2 (β = .668, p < .001) y H3 (efecto indirecto = .176, IC95 % [.096, .259]; efecto directo significativo; VAF = .346) quedaron soportadas, esta última como mediación parcial. H4 no resultó soportada en la ruta DPM → CDL (diferencia = .116, p de permutación = .296), con la limitación derivada de la ausencia de invarianza composicional de CDL; la diferencia entre grupos se localizó en la ruta DPM → CNT (diferencia = −.127, p de permutación = .030).

Figura 13-4

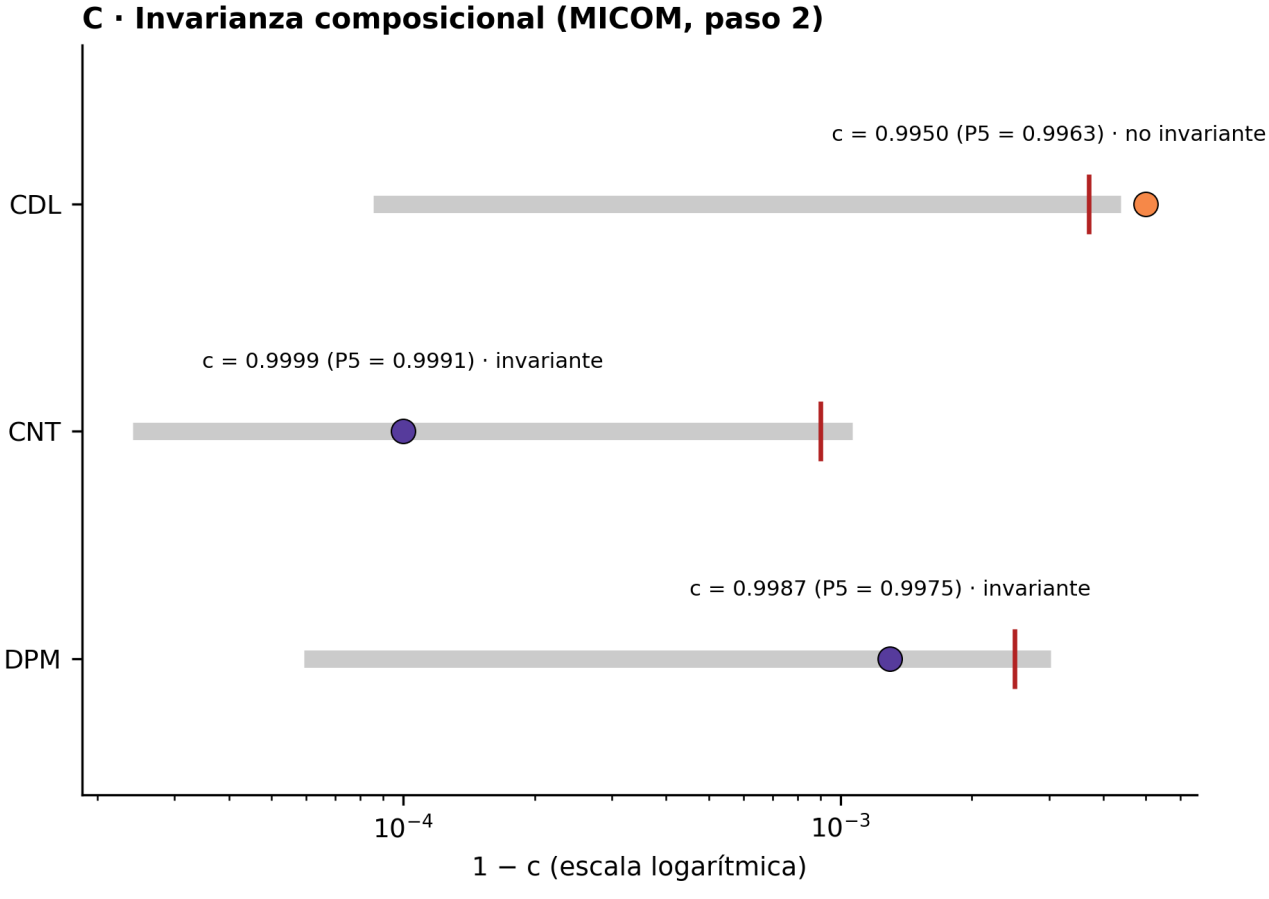
Heterogeneidad entre tipos de institución formadora e invarianza de medición de los compuestos



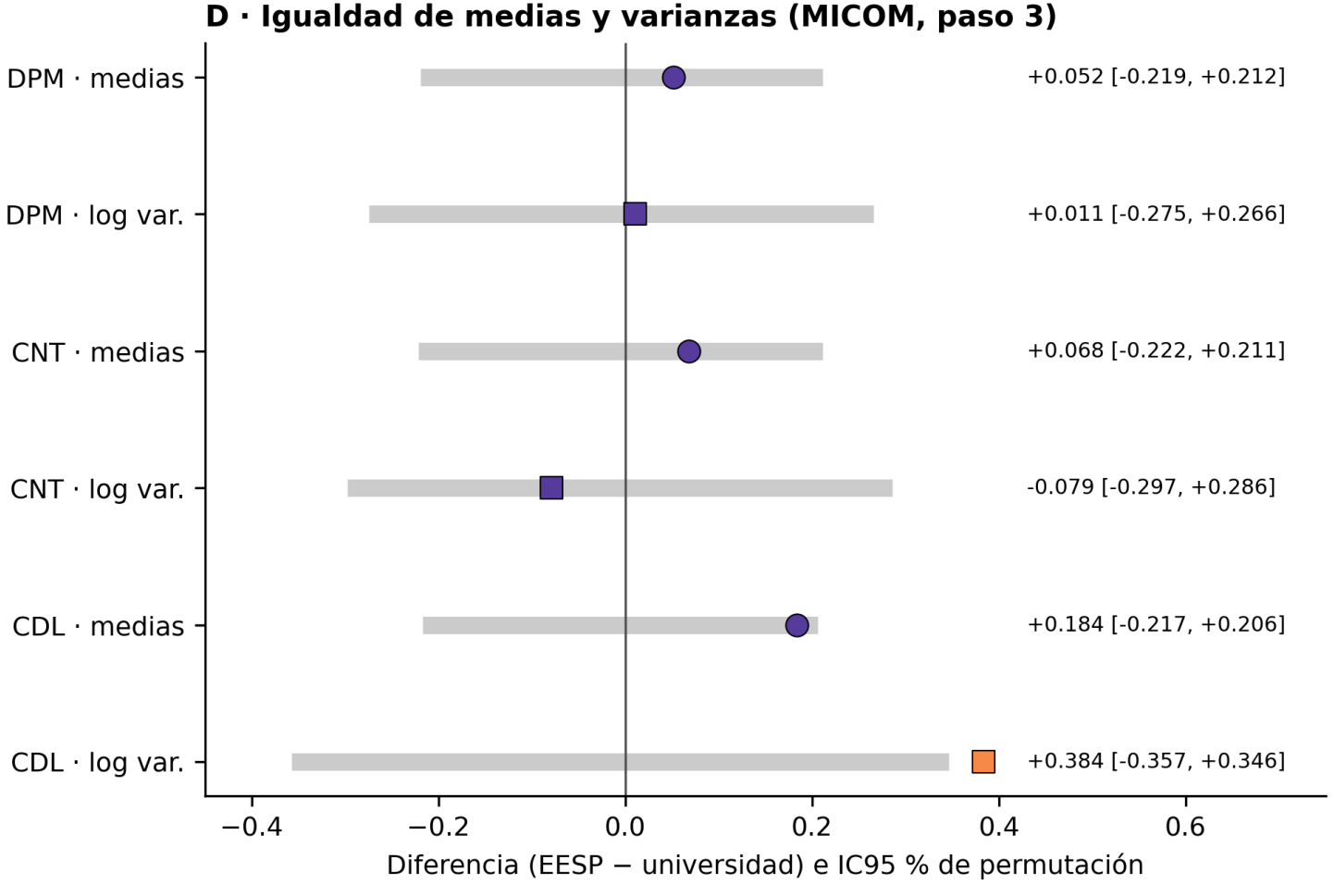
Nota. N = Nota. Universidad n = 179; escuela de educación superior pedagógica n = 167; 5000 permutaciones; valores de las Tablas 13-9 y 13-10. Panel A: coeficientes PLS estimados en cada grupo con IC95 % bootstrap: DPM → CNT .731 frente a .604 (Δ = −.127, p = .030); DPM → CDL .277 frente a .393 (Δ = .116, p = .296); CNT → CDL .243 frente a .302 (Δ = .059, p = .621); efecto indirecto .177 frente a .182 (Δ = .005, p = .957); línea continua para p < .05.



Nota. N = Nota. Universidad n = 179; escuela de educación superior pedagógica n = 167; 5000 permutaciones; valores de las Tablas 13-9 y 13-10. Panel B: distribuciones de permutación de la diferencia en DPM → CNT y DPM → CDL, con la diferencia observada y las colas que definen el p bilateral.



Nota. N = Nota. Universidad n = 179; escuela de educación superior pedagógica n = 167; 5000 permutaciones; valores de las Tablas 13-9 y 13-10. Panel C: paso composicional de MICOM en escala logarítmica de 1 − c: DPM (c = .9987; percentil 5 = .9975) y CNT (c = .9999; percentil 5 = .9991) son invariantes, y CDL (c = .9950; percentil 5 = .9963) no lo es (punto naranja).



Nota. N = Nota. Universidad n = 179; escuela de educación superior pedagógica n = 167; 5000 permutaciones; valores de las Tablas 13-9 y 13-10. Panel D: diferencias de medias y logaritmo de la razón de varianzas con su IC95 % de permutación; todas caen dentro de su intervalo salvo la razón de varianzas de CDL (0.384 [−0.357, 0.346]). Diferencias calculadas como escuela de educación superior pedagógica menos universidad.

13.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre diseño pedagógico y competencia didáctica, sino la disociación entre las dos etapas del mecanismo según la institución: la conversión del diseño pedagógico en competencia narrativa transmedia fue menos intensa en las escuelas de educación superior pedagógica (β = .604) que en las universidades (β = .731) y, sin embargo, el efecto indirecto resultó prácticamente idéntico en ambos grupos (.177 y .182). Si la primera etapa se debilita y la vía mediada no, cabe sospechar que, una vez adquirida, la competencia transmedia rinde de manera similar en la competencia didáctica con independencia de dónde se forme el docente, de modo que el cuello de botella institucional se situaría al comienzo del mecanismo. Esta lectura es conjetural, pues la competencia didáctica literaria no alcanzó invarianza composicional y las rutas que desembocan en ella solo admiten una interpretación orientativa; con todo, la regularidad del efecto indirecto en ambos grupos invita a explorarla con diseños que permitan identificar qué condiciones institucionales intervienen en la etapa inicial.

La asimetría entre las dos rutas que parten del diseño pedagógico (H1, β = .333; H2, β = .668, f^2 = .806) es teóricamente informativa. La literatura sobre didáctica de los géneros breves había documentado que el microrrelato favorece la escritura creativa en plataformas digitales (Mateos y Álvarez, 2021) y que la creación transmedia amplía las alfabetizaciones del estudiantado (Colvert, 2022; García et al., 2024), pero no especificaba por qué vía ese aprendizaje se transfiere a la competencia profesional de quien enseña. La ruta más intensa sugiere que, en la formación docente, el microrrelato opera ante todo como núcleo de expansión multimodal: su condensación, su elipsis y su final abierto invitan a proyectar el relato en otros modos y plataformas, en consonancia con la escritura entendida como diseño semiótico (Moses y Serafini, 2022) y con la brevedad como forma narrativa autónoma que demanda cooperación interpretativa (Roche, 2024).

La mediación parcial (H3) indica que la competencia narrativa transmedia

es una vía real de transferencia, aunque no exclusiva: algo más de un tercio del efecto total discurre por ella, mientras el efecto directo sugiere que el diseño forma competencia didáctica también por mecanismos no mediados, como la modelización de secuencias de lectura y escritura y el uso de criterios explícitos de evaluación literaria (Selfa y Falguera, 2022). Desde una perspectiva crítica, conviene subrayar que el estudio no identifica cuáles son esos mecanismos directos; solo muestra que la vía transmedia no los agota.

En contraste con parte de la literatura reciente, los resultados no apoyan una lectura de la heterogeneidad institucional como diferencia general entre poblaciones de futuros docentes. Los modelos estructurales sobre creencias pedagógicas y adopción tecnológica en docentes en formación describen relaciones no homogéneas (Gurer y Akkaya, 2022), y la evidencia latinoamericana sobre alfabetización transmedia en estudiantes de Comunicación documenta diferencias entre grupos de estudiantes (Rivera y Cruz, 2025). Nuestro patrón es más matizado: la diferencia se concentra en una sola ruta, y precisamente en la única comparación que la medición permite sostener, mientras que la ruta hacia la competencia didáctica no discrimina entre instituciones. Esa precisión es posible porque la comparación se condicionó a la invarianza composicional; los estudios que contrastan grupos sin ese paso previo pueden atribuir al contexto lo que en parte es un artefacto de medición, y el contraste de pendientes sobre compuestos que aquí sugirió una diferencia en la ruta directa, no sostenida bajo el estimador PLS, ilustra ese riesgo. Asimismo, los trabajos que atribuyen la adopción de herramientas digitales para narrativas literarias a la mediación pedagógica, y que registran receptividad docente junto con lagunas y dificultades (Cordeiro et al., 2026), apuntan a condiciones de implementación; nuestros datos son compatibles con esa idea, pero no la contrastan, porque la mediación institucional no fue medida. De igual modo, las experiencias universitarias de integración transmedia, como la de Hiperion (Salazar et al., 2025), y el diagnóstico de las prácticas de lectura literaria en entornos digitales en Ecuador, con sus desafíos y oportunidades (Ortiz et al., 2025), describen contextos concretos sin estimar vías de

transferencia, de modo que la ventaja universitaria observada aquí no puede atribuirse a recursos o infraestructura sin evidencia adicional. Cabe además la posibilidad de que la exposición diferencial a proyectos de innovación o a la formación disciplinar en las universidades explique la mayor conversión inicial, hipótesis que el diseño transversal no permite someter a prueba.

La falta de invarianza de la competencia didáctica literaria merece una lectura propia. Los estudios sobre formación inicial docente la presentan como un conjunto heterogéneo de capacidades: la mediación entre el alumnado y el texto literario mediante el videopoema (Ibarra et al., 2026), la selección de obras multimodales como competencia de mediación (Santiago et al., 2026), la reflexión sobre la lectura literaria en la formación en Letras (Sardi, 2023) y la articulación entre lectura y creación (Contreras, 2023). Es plausible que facetas como seleccionar, interpretar, articular lectura con creación y evaluar se ponderen de manera distinta en el currículo de cada tipo de institución, lo que alteraría los pesos con que el compuesto se forma en cada grupo. Esta explicación es tentativa y sugiere, más que prueba, que futuros trabajos conciban la competencia didáctica como compuesto de orden superior o la midan con instrumentos que preserven la comparabilidad entre contextos formativos.

Los resultados coinciden, en cambio, con la evidencia de que las prácticas narrativas digitales y multimodales desarrollan capacidades profesionales en el profesorado en formación: el relato digital fortalece el pensamiento crítico (Isaacs et al., 2024), la composición de narrativas gráficas moviliza convenciones multimodales (Cook y Chisholm, 2025) y el lector multimodal ocupa un lugar creciente en la educación literaria (Ibarra y Ballester, 2025). Del mismo modo, el patrón de mediación parcial es coherente con la transmisión de los efectos de la enseñanza a través de capacidades del propio docente (Hu y Choi, 2023; Hodges et al., 2021), si bien aquí el mediador no es la autoeficacia, sino una competencia narrativa específica de entornos transmedia.

La contribución trasciende el caso peruano, que funciona como sitio de prueba por la coexistencia de dos tradiciones institucionales de formación docente y de una tradición propia de microrrelato (Llerena, 2023). En el plano teórico, el estudio especifica y contrasta un mecanismo de transferencia —del diseño pedagógico de un género breve a la competencia didáctica literaria a través de la competencia narrativa transmedia— que responde a la carencia de modelos cuantitativos señalada por las revisiones sobre narrativa transmedia en educación (Meyerhofer, 2025; Meyerhofer y González, 2025) y que puede contrastarse en cualquier sistema que trabaje con géneros breves en entornos digitales. En el plano metodológico, muestra que la heterogeneidad entre instituciones solo es interpretable cuando se condiciona a la invarianza composicional y se contrasta con permutación sobre el propio estimador.

El modelo explicó el 44.6 % de la varianza de la competencia narrativa transmedia y el 29.7 % de la competencia didáctica literaria, con capacidad predictiva fuera de muestra y un error inferior al del modelo lineal de referencia (Hair et al., 2021; Sharma et al., 2023). Que algo más de dos tercios de la varianza de la competencia didáctica quede sin explicar es esperable en un modelo parsimonioso con un solo predictor exógeno, e indica que otras condiciones —la formación disciplinar en literatura, la práctica preprofesional o la experiencia lectora— completan la explicación.

13.6 CONCLUSIONES

El aporte teórico principal es que la competencia narrativa transmedia constituye una vía de transferencia, parcial pero sustantiva, por la que el diseño pedagógico del microrrelato se convierte en competencia didáctica literaria, lo que articula la teoría de la multimodalidad y la alfabetización transmedia en un mecanismo contrastado empíricamente. La heterogeneidad entre tipos de institución se localiza en la primera etapa de ese mecanismo y no en la relación directa con la competencia didáctica.

Las implicaciones prácticas son tres:

1) Conviene incorporar de manera sistemática tareas de expansión transmedia del microrrelato —adaptación a formatos visuales y audiovisuales, recorridos narrativos entre plataformas y participación de los lectores— en la formación inicial, dado el peso de la ruta del diseño pedagógico hacia la competencia transmedia ($\beta = .668$).

2) Esas tareas deben acompañarse con criterios explícitos de evaluación literaria, pues el efecto directo del diseño pedagógico ($c' = .333$) no depende solo de la vía transmedia.

3) Resulta recomendable reforzar en las escuelas de educación superior pedagógica las condiciones que convierten el diseño pedagógico en competencia transmedia, donde esa conversión fue menor ($\beta = .604$ frente a .731).

Las limitaciones son las siguientes: el diseño transversal no permite establecer la dirección causal de las relaciones; la selección por conveniencia restringe la generalización de las estimaciones a la población de docentes en formación; las medidas son de autoinforme, aunque los factores de inflación de la varianza de colinealidad completa (máximo 2.496) no indicaron sesgo de método común; y la ausencia de invarianza composicional de la competencia didáctica literaria restringe la comparación multigrupo de las rutas que desembocan en ella, por lo que el no soporte de H4 no equivale a equivalencia entre instituciones.

Quedan abiertas varias preguntas. ¿Qué condiciones institucionales —currículo, práctica preprofesional, acceso a recursos multimedia— explican la diferencia en la ruta hacia la competencia narrativa transmedia entre universidades y escuelas de educación superior pedagógica? ¿Media la competencia narrativa transmedia relaciones análogas en otros géneros breves, como la poesía digital? ¿Confirmaría un diseño longitudinal o cuasiexperimental la dirección del mecanismo propuesto?

13.7 REFERENCIAS BIBLIOGRÁFICAS

Allué, C., & Cassany, D. (2023). Grabando vídeos: educación literaria multimodal. *Texto Livre, 16*. https://doi.org/10.1590/1983-3652.2023.41797

Bartolomé, C. (2023). Definición y canon del microrrelato en los manuales españoles de Secundaria y Bachillerato. *Microtextualidades. Revista Internacional de microrrelato y minificción*(13), 68-89. https://doi.org/10.31921/microtextualidades.n13a5

Chin, W. (1998). The partial least squares approach to structural equation modeling. En G. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Lawrence Erlbaum Associates

Chuquilin, J. (2026). Políticas de formación inicial docente en Bolivia y Perú: Entre la descolonización y la tecnocracia. *Revista Mexicana de Investigación Educativa, 31*(110). https://doi.org/10.66708/rmie.v31i110.2040

Clavijo, Á. (2026). La relectura de la cuentística tradicional: un enfoque transmedia para la educación literaria en la Educación Superior. *Ocnos, 25*(2). https://doi.org/10.18239/ocnos_2026.25.2.625

Colvert, A. (2022). Dreams of time and space: exploring digital literacies through playful transmedia storying in school. *Literacy, 56*(1), 59-72. https://doi.org/10.1111/lit.12271

Contreras, E. (2023). Lectura y creación: Un estudio sobre lectura literaria y creatividad en formación inicial de docentes. *Ocnos, 22*(2). https://doi.org/10.18239/ocnos_2023.22.2.383

Cook, M., & Chisholm, J. (2025). Composing comics as activism: Leveraging preservice teachers' multimodal and graphic narrative conventions. *Journal of Adolescent & Adult Literacy, 68*(6), 690-704. https://doi.org/10.1002/jaal.70002

Cordeiro, F., Da Silva, A., & Damasceno, E. (2026). Pedagogical mediation drives teacher adoption of visual novel for literary narratives. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1753522

Doria, D. (2023). Reflexiones sobre la palabra en un libro de minificción de Westphalen. *Microtextualidades. Revista Internacional de microrrelato y minificción*(13), 15-24. https://doi.org/10.31921/microtextualidades.n13a2

Farfán, Z. (2026). Alfabetizaciones múltiples, narrativas transmedia e interactividad: un acercamiento al estado de la cuestión. *Revista Educación*. https://doi.org/10.15517/revedu.v50i2.1069

Fernández, A., Zermeño, A., & Cornelio, E. (2025). El uso de una narrativa transmedia como eje central de una estrategia didáctica multidisciplinaria. *Revista Iberoamericana de Educación Superior*, 101-121. https://doi.org/10.22201/iisue.20072872e.2025.46.1777

García, M. (2025). Literatura y transmedia. Educación literaria en el nivel medio superior. *Revista RedCA, 7*(21), 209. https://doi.org/10.36677/redca.v7i21.24082

García, M., Mañas, M., Aranda, L., & Cortés, P. (2024). Educational transmedia narratives and the INAEP method: innovation and

learning in the Colombian Context. Pilot study. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1406666

Gurer, M., & Akkaya, R. (2022). The influence of pedagogical beliefs on technology acceptance: a structural equation modeling study of pre-service mathematics teachers. *Journal of Mathematics Teacher Education, 25*(4), 479-495. https://doi.org/10.1007/s10857-021-09504-5

Hair, J., Hult, G., Ringle, C., Sarstedt, M., Danks, N., & Ray, S. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R*. Springer. https://doi.org/10.1007/978-3-030-80519-7

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405-431. https://doi.org/10.1108/IMR-09-2014-0304

Hodges, T., Wright, K., & McTigue, E. (2021). The Preservice Teacher Self-Efficacy for Writing Inventory (PTSWI): A tool for measuring beliefs about writing. *Assessing Writing, 49*, 100545. https://doi.org/10.1016/j.asw.2021.100545

Hu, Y., & Choi, J. (2023). Is creative teaching behaviours' effect on English creative writing multi-mediated by writing and creative

self-efficacy?. *New Writing*, *20*(4), 383-399. https://doi.org/10.1080/14790726.2023.2188223

Ibarra, N., & Ballester, J. (2025). Cómic y educación literaria: Del utilitarismo y la controversia epistemológica al lector multimodal. *Bellaterra: Journal of Teaching and Learning Language and Literature*, *18*(4), e1583. https://doi.org/10.5565/rev/jtl3.1583

Ibarra, N., Ballester, J., & Rodrigo, F. (2026). Multimodalidad y educación literaria: el videopoema en la formación de profesorado. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, *29*(2), 119-135. https://doi.org/10.6018/reifop.716391

Isaacs, M., Tondeur, J., & Vaesen, J. (2024). Digital storytelling in teacher education: Developing pre-service teachers' critical thinking. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.9258

Isshiki, T. (2026). Validation of the creative educational practices scale: Measuring teachers' everyday creativity. *Teaching and Teacher Education, 181*, 105693. https://doi.org/10.1016/j.tate.2026.105693

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.

Jordan, A. (2023). Formación docente en competencias digitales para el área de lengua y literatura. *Mérito - Revista de Educación*, *5*(14), 49-60. https://doi.org/10.33996/merito.v5i14.1122

Klesel, M., Schuberth, F., Niehaves, B., & Henseler, J. (2022). Multigroup Analysis in Information Systems Research using PLS-PM: A Systematic Investigation of Approaches. *ACM SIGMIS Database: the*

DATABASE for Advances in Information Systems, 53(3), 26-48. https://doi.org/10.1145/3551783.3551787

Kock, N. (2015). Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach. *International Journal of e-Collaboration, 11*(4), 1-10. https://doi.org/10.4018/ijec.2015100101

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227-261. https://doi.org/10.1111/isj.12131

Llerena, P. (2023). Las antologías de microrrelato peruanas: apuntes sobre su estudio y conformación. *Microtextualidades. Revista Internacional de microrrelato y minificción*(13), 1-14. https://doi.org/10.31921/microtextualidades.n13a1

Mackey, M. (2025). Novice interpreters, transmedia fictions and the afferent stance. *Literacy, 59*(2), 177-184. https://doi.org/10.1111/lit.12381

Mateos, B., & Álvarez, E. (2021). Los microrrelatos hipermediales de Patricia Esteban Erlés: didáctica de la fantasía. *Microtextualidades. Revista Internacional de microrrelato y minificción*(10), 35-56. https://doi.org/10.31921/microtextualidades.n10a3

Meyerhofer, R. (2025). A systematic review of transmedia literacy: Implications for lifelong learning. *Journal of Media Literacy Education, 17*(2), 110-124. https://doi.org/10.23860/jmle-2025-17-2-8

Meyerhofer, R., & González, J. (2025). So, what do we talk about when we talk about transmedia storytelling in education? A systematic literature review. *Review of Education, 13*(2). https://doi.org/10.1002/rev3.70086

Meyerhofer, R., González, J., & Peracaula, M. (2024). Postdigital Storytelling: Storytelling (Within or Across) the Digital and Transmedia Field. *Postdigital Science and Education, 6*(3), 886-901. https://doi.org/10.1007/s42438-024-00476-2

Moses, L., & Serafini, F. (2022). Image, text and design: students' semiotic choices in nonfiction compositions. *Literacy, 56*(4), 340-354. https://doi.org/10.1111/lit.12293

Nabhan, S., & Habók, A. (2025). Development and validation of the Teacher Multiliteracies Scale (TMS) for English language teacher education. *System, 131*, 103673. https://doi.org/10.1016/j.system.2025.103673

Ortiz, J., Mora, L., & Aldaz, A. (2025). Prácticas de lectura literaria en entornos digitales: desafíos y oportunidades para la educación literaria en Ecuador. *RICEd: Revista de Investigación en Ciencias de la Educación, 3*(6), 32-43. https://doi.org/10.53877/ms268561

Pérez, A. (2023). Una perspectiva crítico-social sobre la concepción de la alfabetización transmedia. *Educación y Ciencia, 27*. https://doi.org/10.19053/0120-7105.eyc.2023.27.e16276

Rivera, B. (2024). Diseño, validación y fiabilidad de un instrumento para la medición de la alfabetización transmedia en estudiantes de comunicación. *Emerging Trends in Education, 7*(13), 75-88. https://doi.org/10.19136/etie.a7n13.6046

Rivera, B., & Cruz, L. (2025). Brecha de género en Alfabetización Transmedia de estudiantes de Comunicación del Centro Occidente de México. *Revista Convergencia Educativa*(18), 1-18. https://doi.org/10.29035/rce.18.1

Roche, S. (2024). Flash fiction as a distinct literary form: some thoughts on time, space, and context. *New Writing, 21*(2), 171-189. https://doi.org/10.1080/14790726.2023.2293767

Salazar, M., Sepúlveda, M., & Álvarez, J. (2025). Hiperion: integrando las narrativas transmedia en los contextos universitarios. *Comunicación*(52), 1-25. https://doi.org/10.18566/comunica.n52.a04

Santiago, M., Rivera, P., Alcántara, L., & Sánchez, S. (2026). La selección de álbumes ilustrados como competencia para la mediación en lectura literaria en la formación inicial docente. *Revista Electrónica Interuniversitaria de Formación del Profesorado, 29*(2), 1-15. https://doi.org/10.6018/reifop.715681

Sardi, V. (2023). Para una erótica de la lectura literaria en la formación docente en Letras. *Revista del IICE*(54). https://doi.org/10.34096/iice.n54.14047

Selfa, M., & Falguera, E. (2022). From Text on Paper to Digital Poetry: Creativity and Digital Literary Reading Practices in Initial Teacher Education. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.882898

Sharma, P., Liengaard, B., Hair, J., Sarstedt, M., & Ringle, C. (2023). Predictive model assessment and selection in composite-based

modeling using PLS-SEM: extensions and guidelines for using CVPAT. *European Journal of Marketing*, *57*(6), 1662-1677. https://doi.org/10.1108/EJM-08-2020-0636

Sánchez, S. (2025). De la reflexión gramatical a la comprensión textual, a propósito de un microrrelato. *Huarte de San Juan. Filología y Didáctica de la Lengua*(24), 65-89. https://doi.org/10.48035/rhsj-fd.24.3

Tombleson, B. (2024). Transmedia learning: a literature review. *Technology, Pedagogy and Education*, *33*(2), 255-269. https://doi.org/10.1080/1475939x.2024.2310681

CAPÍTULO XIV

TERTULIAS LITERARIAS DIALÓGICAS Y LECTURA CRÍTICA DE LA NARRATIVA INDIGENISTA EN LA EDUCACIÓN SECUNDARIA PERUANA: UN MODELO DE ECUACIONES ESTRUCTURALES LATENTE DE DOS OLEADAS CON LA EMPATÍA LITERARIA INTERCULTURAL COMO MECANISMO

Eugenio Ccanto Mallma

Universidad Nacional Intercultural de la Amazonía

Resumen

Se estimó el efecto de las tertulias literarias dialógicas con narrativa indigenista peruana sobre la lectura crítica en secundaria y se contrastó si opera en parte mediante la empatía literaria intercultural. Cuasiexperimento pretest-postest con comparación activa en 96 aulas de 48 instituciones públicas de Lima Metropolitana, Ayacucho y Cusco (2025): en cada institución, un aula desarrolló catorce tertulias semanales y otra trabajó los mismos textos con la enseñanza habitual. Se analizaron 2 422 estudiantes; la muestra se fijó a priori por el efecto indirecto. Ambos constructos se modelaron como variables latentes con invarianza escalar, en un modelo de ecuaciones estructurales de dos oleadas con bootstrap por institución. Las tertulias aumentaron la empatía (d = 0.17) y la lectura crítica latente (efecto total d = 0.25, IC 95 % [0.16, 0.33]). La empatía postest predijo la lectura crítica controlando ambos pretests (β = .190); el efecto indirecto fue d = 0.032 (IC 95 % [0.019, 0.045]), 12.9 % del total. El efecto resistió cuatro especificaciones alternativas y no difirió según lengua materna quechua ni ámbito rural. Discutir en diálogo igualitario literatura que representa el mundo andino mejora la lectura crítica, y una parte modesta de esa mejora pasa por comprender a personajes de otra cultura.

Palabras clave: tertulias literarias dialógicas; lectura crítica; narrativa indigenista; empatía literaria; educación intercultural; modelos de ecuaciones estructurales.

Abstract

We estimated the effect of dialogic literary gatherings with Peruvian Indigenist narrative on critical reading in secondary school and tested whether it operates partly through intercultural literary empathy. A pretest-posttest quasi-experiment with an active comparison group was conducted in 96 classrooms of 48 public schools in Metropolitan Lima, Ayacucho and Cusco (2025): in each school, one classroom held fourteen weekly gatherings and another worked on the same

texts with regular instruction. We analysed 2,422 students; sample size was set a priori for the indirect effect. Both constructs were modelled as latent variables with scalar invariance, in a two-wave structural equation model with school-level bootstrap. The gatherings increased empathy (d = 0.17) and latent critical reading (total effect d = 0.25, 95% CI [0.16, 0.33]). Posttest empathy predicted critical reading controlling for both pretests (β = .190); the indirect effect was d = 0.032 (95% CI [0.019, 0.045]), 12.9% of the total. The effect held across four alternative specifications and did not differ by Quechua home language or rural setting. Discussing literature that represents the Andean world under egalitarian dialogue improves critical reading, and a modest share of that improvement runs through understanding characters from another culture.

Keywords: *dialogic literary gatherings; critical reading; Indigenist narrative; literary empathy; intercultural education; structural equation modeling.*

14.1 INTRODUCCIÓN

Leer de manera crítica un texto literario exige más que reconstruir lo que dice: supone inferir motivos, reconocer qué voces tienen poder dentro del relato y sostener un juicio propio con evidencias (Baki, 2025; Coleman, 2025). La investigación reciente ha documentado que las estrategias de lectura crítica pueden enseñarse, que su instrucción mejora la comprensión y la motivación lectora (Gebremariam y Weldeyohannes, 2025) y la habilidad de lectura crítica en aulas de lengua extranjera (Mekuria et al., 2024), y que la lectura crítica se asocia con el pensamiento crítico (Baki, 2025). El campo de la literacidad crítica advierte, sin embargo, que reducir lo crítico a la sospecha frente al texto empobrece las formas de leer el poder que la escuela puede cultivar (Coleman, 2025).

Las síntesis disponibles sobre la enseñanza de la comprensión se concentran en programas dialógicos para la educación infantil y primaria (Xie y Lin, 2025) y en combinaciones de estrategias para lectores con dificultades de tercero a duodécimo grado (Peng et al., 2024). Queda menos explorada la lectura crítica de textos literarios en secundaria, donde la discusión puede ser el vehículo principal del aprendizaje.

Las tertulias literarias dialógicas son una actuación educativa en la que estudiantes de todas las edades leen obras de la literatura clásica y las discuten en condiciones de diálogo igualitario: cada participante elige un fragmento, lo comparte y argumenta su sentido. Una revisión sistemática de 46 estudios publicados entre 2010 y 2021 concluyó que las tertulias favorecen el intercambio, la superación de estereotipos y barreras culturales y la mejora de las habilidades lingüísticas y lectoras en contextos muy diversos (Ruiz et al., 2023). La evidencia posterior se ha ampliado a escuelas de educación especial (Fernández et al., 2021; Roca et al., 2024), a aprendices de una segunda lengua (Santiago et al., 2023; Fernández y García, 2025), a estudiantes marginados de Ghana (Allotey et al., 2023), a centros de acogida (Salceda et al., 2022) y a la prevención del acoso escolar (Álvarez et al., 2023).

Ese cuerpo de evidencia tiene, sin embargo, una forma característica: casi todo procede de estudios de caso y de métodos cualitativos o comunicativos, con grupos pequeños y medidas de percepción (Khaqan et al., 2025; Fernández et al., 2025). Faltan estimaciones del tamaño del efecto con grupo de comparación, con medición cuya equivalencia entre oleadas y condiciones esté contrastada y, sobre todo, pruebas del mecanismo por el que el diálogo sobre la literatura mejoraría la lectura. Los trabajos recientes que vinculan la interacción dialógica con el rendimiento lector lo hacen con diseños correlacionales (Hogan y Sun, 2025; Tong y Ding, 2025) o con comparaciones entre aproximaciones de enseñanza de la literatura sin modelo de mecanismo (Fulton et al., 2023).

Este vacío adquiere consecuencias propias en el contexto peruano. La educación secundaria combina una brecha persistente en lectura, documentada en las evaluaciones internacionales de 2022 (OECD, 2023; Arias et al., 2023), con una diversidad lingüística y cultural que la política de educación intercultural bilingüe reconoce pero que llega con dificultad a las aulas (Liñán et al., 2023; Sullón, 2024). Las brechas lectoras en contextos rurales y desfavorecidos son persistentes (Kiziltaş y Kultas, 2026), y los estudiantes de lengua materna originaria encuentran pocas oportunidades de ver su mundo representado en el currículo (Marino et

al., 2023). La narrativa indigenista peruana —de Enrique López Albújar y Ciro Alegría a José María Arguedas y Manuel Scorza— representa ese mundo andino, sus conflictos de poder y sus lenguas, y forma parte del currículo de Comunicación; pero suele enseñarse mediante cuestionarios de comprensión que dejan poco espacio a la interpretación propia.

Una segunda línea de investigación sugiere por qué la discusión literaria podría beneficiar la lectura crítica. La evidencia sobre la ficción y la cognición social no es unánime (Rezende y Shigaeff, 2023), pero en la adolescencia la empatía con los personajes parece transferirse a personas reales y vincularse con la trayectoria lectora (Santi et al., 2025; Lenhart et al., 2022). Cuando los personajes pertenecen a una cultura históricamente subordinada, esa empatía adquiere una dimensión intercultural: comprender por qué actúan como actúan comuneros quechuas frente al gamonalismo exige suspender el estereotipo y reconocer otra lógica (Zsögön, 2025). El aprendizaje intercultural se ha vinculado con la competencia global de los adolescentes a través de mecanismos psicológicos intermedios (Huang et al., 2025), y las prácticas más eficaces combinan aprender sobre otra cultura, reflexionar sobre ella y experimentarla (Al-Afifi et al., 2026).

El aporte del estudio no es regional. Perú funciona como un caso de prueba exigente para una pregunta general de la investigación en lectura y educación intercultural: si la comprensión empática de personajes de otra cultura es un mecanismo por el que la discusión literaria mejora la lectura crítica. Responderla requiere tres condiciones que la literatura previa no ha reunido: un contraste con grupo de comparación que trabaje los mismos textos, medidas latentes cuya equivalencia se demuestre antes de comparar y una muestra suficiente para estimar un efecto indirecto en datos agrupados por aula. Este trabajo las reúne.

14.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

14.2.1 Diálogo igualitario y lectura crítica

La enseñanza dialógica sostiene que el aprendizaje se profundiza cuando los estudiantes co-construyen significados, expresan interpretaciones múltiples y las justifican ante otros. En la lectura de textos, ello requiere posturas lectoras que abran el texto a la discusión; las posturas críticas o expresivas pueden incluso estrechar las interpretaciones si no se acompañan de un diálogo sostenido (Barak y Lefstein, 2022). Las síntesis cuantitativas informan efectos grandes de la lectura dialógica y de los círculos literarios (Tang et al., 2026; Ma, 2025) y efectos pequeños de la formación docente en comprensión (Rice et al., 2024). En la tertulia, cada estudiante elige y argumenta un fragmento; esa estructura convierte la interpretación propia, el análisis de las voces del relato y la evaluación argumentada en el núcleo de la actividad, que son las tres dimensiones de lectura crítica aquí medidas.

- **H1.** Las tertulias literarias dialógicas con narrativa indigenista mejoran la lectura crítica, controlando el nivel inicial, frente a la enseñanza habitual de los mismos textos.

14.2.2 La empatía literaria intercultural como mecanismo

Leer ficción ofrece la posibilidad de habitar la perspectiva de personajes distintos de uno mismo, con una mejora pequeña pero robusta de la cognición social (Dodell y Tamir, 2018). La discusión parece ser el ingrediente que transforma la lectura en comprensión de otros: discutir ficción literaria en pequeños grupos mejora la toma de perspectiva social (Frei et al., 2026). Cuando los personajes pertenecen a una cultura subordinada, esta comprensión adquiere carácter intercultural (Walsh, 2023). A su vez, es condición de la lectura crítica de un texto indigenista: no se puede analizar qué voces tienen poder en el relato ni evaluar la justicia de lo narrado sin reconstruir antes la perspectiva de quienes lo viven.

- **H2.** Las tertulias aumentan la empatía literaria intercultural.
- **H3**. La empatía literaria intercultural al final de la intervención predice la lectura crítica final, controlando ambos niveles iniciales.
- **H4**. El efecto de las tertulias sobre la lectura crítica está mediado parcialmente por la empatía literaria intercultural.

14.2.3 Equidad del efecto

Si las tertulias con narrativa indigenista funcionan como una pedagogía que reconoce la experiencia de los estudiantes de lengua originaria y de ámbito rural, cabría esperar un efecto al menos igual para ellos. Se plantea, por ello, como pregunta de investigación si el efecto difiere según la lengua materna quechua y el ámbito rural de la institución.

14.3 MATERIALES Y MÉTODOS

14.3.1 Diseño y participantes

Se empleó un cuasiexperimento pretest-postest con grupo de comparación activo y asignación por aula dentro de cada institución. En 48 instituciones públicas, dieciséis por región en Lima Metropolitana, Ayacucho y Cusco, se incluyeron dos aulas de cuarto o quinto grado de secundaria; el equipo directivo decidió cuál desarrollaría las tertulias y cuál trabajaría los mismos textos con la enseñanza habitual, por lo que la asignación no fue probabilística. El pretest se aplicó en marzo y el postest en agosto de 2025. Las instituciones se incorporaron por disponibilidad, sin pretensión de representatividad poblacional. Se matricularon 2 606 estudiantes en 96 aulas y 2 422 completaron ambas mediciones (1 196 en tertulias y 1 226 en comparación).

14.3.2 Tamaño de la muestra

El tamaño se determinó a priori con tres criterios, y la N del estudio es la que exige el más demandante (Tabla 14-1). Para el efecto total (H1)

se calculó la potencia de la prueba t no central del análisis de covarianza con asignación por aula, con $\delta = 0.30$, $\rho = .20$ como valor de planificación conservador (Hedges y Hedberg, 2007) y 25 estudiantes por aula. Para el efecto indirecto (H4) se usó el método de Monte Carlo para modelos de mediación (Schoemann et al., 2017; Wang y Rhemtulla, 2021), con 600 réplicas del diseño agrupado. Para el ajuste del modelo se empleó la prueba de ajuste cercano basada en el RMSEA (MacCallum et al., 1996). La N determinada es la del efecto indirecto, 2 400 estudiantes en 96 aulas; previendo una pérdida cercana al 7 %, la muestra analizada la cubre, conforme a las recomendaciones de potencia para modelos estructurales (Moshagen y Bader, 2023).

Tabla 14-1

Cálculo a priori del tamaño de la muestra para los tres contrastes del estudio ($\delta = .05$; potencia objetivo = .80)

Contraste		Método	Supuestos de cálculo	Resultado	N mínima
H1:	efecto total sobre la lectura crítica	Potencia de la prueba t no central del análisis de covarianza con asignación por aula	$\delta = 0.30$; $\rho = .20$; R^2 del pretest = .36 (estudiante) y .50 (aula); n = 25 por aula	46 aulas: potencia = .817 (λ = 2.93; gl = 43)	1 150
H4:	efecto indirecto a × b	Monte Carlo con 600 réplicas e intervalo de Monte Carlo de 4 000 extracciones, errores robustos por aula	a = 0.20 DE; b = .20; c′ = 0.25; CCI residual = .10	72 aulas: potencia = .730 (a: .703; b: 1.000)	—
		Monte Carlo con 600 réplicas e intervalo de Monte Carlo de 4 000 extracciones, errores robustos por aula	a = 0.20 DE; b = .20; c′ = 0.25; CCI residual = .10	88 aulas: potencia = .760 (a: .747; b: 1.000)	—
		Monte Carlo con 600 réplicas e intervalo de Monte Carlo de 4 000 extracciones, errores robustos por aula	a = 0.20 DE; b = .20; c′ = 0.25; CCI residual = .10	96 aulas: potencia = .818 (a: .810; b: 1.000)	2 400

Contraste	Método	Supuestos de cálculo	Resultado	N mínima
Ajuste aproximado del modelo estructural	Prueba de ajuste cercano basada en el RMSEA (MacCallum et al., 1996)	gl = 316; ε_0 = .05; ε = .08	N = 64: potencia = .803	64
Muestra del estudio	El contraste más exigente determina la N	96 aulas de 48 instituciones educativas	Analizados: 2 422 estudiantes (matriculados: 2 606)	2 400

Nota: δ = efecto total en desviaciones estándar; ρ = correlación intraclase no condicional; CCI residual = correlación intraclase tras ajustar por el pretest. Las tres filas del contraste H4 muestran la trayectoria de la potencia hasta alcanzar el objetivo..

14.3.3 Intervención e instrumentos

En las aulas experimentales, el docente de Comunicación desarrolló catorce tertulias semanales de 90 minutos entre abril y julio de 2025. Cada estudiante marcaba un pasaje del fragmento acordado, lo leía y explicaba por qué lo había elegido, y el docente moderaba garantizando que las intervenciones se valoraran por sus argumentos y no por quién las emitía, sin ofrecer una interpretación correcta de antemano (Girbés et al., 2024). El corpus incluyó textos de López Albújar, Arguedas, Alegría y Scorza. Las aulas de comparación leyeron las mismas obras con lectura guiada y fichas de comprensión. La asistencia media fue de 12.48 sesiones (DE = 1.28; rango 7–14).

La lectura crítica se midió con seis tareas de respuesta construida (0–10, rúbrica analítica) sobre un fragmento indigenista no trabajado en clase, con formas paralelas, agrupadas en tres dimensiones; dos evaluadores ciegos a la condición corrigieron de forma independiente el 20 % de las respuestas. La empatía literaria intercultural se midió con cinco ítems Likert, dos inversos (Tabla 14-2).

Tabla 14-2

Operacionalización de los constructos, descriptivos por ítem y fiabilidad entre evaluadores (N = 2 422)

Constructo y dimensión	Código	Indicador	Escala	M (DE)		CCI entre evaluadores (pre / post)
				pretest	postest	
Lectura crítica — Interpretación inferencial	T1	Explica, con evidencia del texto, por qué el personaje actúa como actúa ante el conflicto central	0–10 (rúbrica analítica)	5.58 (1.89)	5.86 (1.98)	.918 / .925
	T2	Infiere el sentido de un símbolo o imagen andina recurrente y lo relaciona con el desenlace		5.16 (1.91)	5.45 (1.96)	.923 / .922
Lectura crítica — Análisis ideológico-intercultural	T3	Identifica qué voces (comuneros, hacendados, narrador) tienen poder en el relato y cómo lo muestra el texto		4.56 (1.94)	4.88 (2.00)	.923 / .922
	T4	Contrasta la representación del mundo andino del relato con una visión estereotipada y argumenta la diferencia		4.37 (1.89)	4.70 (1.96)	.928 / .911
Lectura crítica — Evaluación argumentada	T5	Formula y defiende un juicio sobre la justicia de la acción narrada, con dos evidencias textuales		4.10 (1.93)	4.39 (1.99)	.924 / .910
	T6	Responde a un argumento contrario al propio y lo refuta o lo integra con razones		3.87 (1.90)	4.19 (1.95)	.913 / .923
Empatía literaria intercultural	—	Cuando leo, entiendo por qué los personajes de las comunidades andinas sienten lo que sienten	1–5 (Likert)	3.02 (1.07)	3.13 (1.07)	—
	—	Las historias de personajes quechuas me ayudan a ver el mundo desde su lugar		3.07 (1.06)	3.19 (1.07)	—
	—	Me afecta lo que les pasa a los personajes aunque vivan de un modo muy distinto al mío		3.14 (1.08)	3.22 (1.07)	—
	—	Me cuesta ponerme en el lugar de personajes de otra cultura (ítem inverso)		3.17 (1.07)	3.30 (1.09)	—
	—	Las costumbres de los personajes andinos me parecen ajenas a mi vida (ítem inverso)		3.26 (1.06)	3.34 (1.08)	—

Nota: CCI = coeficiente de correlación intraclase de acuerdo absoluto para una sola medida, calculado sobre las 500 respuestas corregidas de forma independiente por dos evaluadores. Los ítems inversos se recodificaron antes del análisis. Las tareas del pretest y del postest son formas paralelas con fragmentos narrativos distintos de la misma tradición literaria.

14.3.4 Procedimiento y análisis

Personal externo aplicó las pruebas en horario de clase (80 minutos la prueba de lectura y 10 la escala). Los docentes experimentales recibieron formación inicial y acompañamiento quincenal.

El análisis siguió cuatro pasos:

1) Se evaluó la equivalencia inicial con diferencias estandarizadas y la pérdida entre oleadas con regresión logística con errores robustos por institución.

2) Se estimó un análisis factorial confirmatorio de cuatro factores y se contrastó la invarianza configural, métrica y escalar, longitudinal y entre condiciones, mediante la diferencia de χ^2 entre modelos anidados (Luong y Flake, 2023; Mackinnon et al., 2022); la validez discriminante se evaluó con el criterio de Fornell y Larcker (1981) y la razón heterorrasgo-monorrasgo (Henseler et al., 2015).

3) Se estimó un modelo de ecuaciones estructurales latente de dos oleadas en forma de análisis de covarianza, con cargas iguales entre oleadas, la condición como predictor de la empatía y de la lectura crítica postest y ambos pretests como control, además de covariables sociodemográficas (Tabla 14-8); el ajuste se evaluó con CFI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu y Bentler, 1999). Los efectos indirecto, directo y total se estimaron con intervalos percentiles de 1 000 remuestreos de instituciones (Falk et al., 2023; Tibbe y Montoya, 2022) y se expresaron en desviaciones estándar de la variable dependiente (Flora et al., 2025).

4) Se contrastó la robustez con modelos multinivel, de ganancia, de ponderación por propensión y de efectos fijos de institución, y se exploró la heterogeneidad con interacciones por lengua materna y ámbito. Se usó Python 3.12 con semopy, statsmodels y SciPy; las respuestas de cinco puntos se trataron como continuas.

14.4 RESULTADOS

14.4.1 Equivalencia inicial y pérdida entre oleadas

Las dos condiciones fueron comparables al inicio: todas las diferencias estandarizadas fueron pequeñas, y la mayor en valor absoluto correspondió a la empatía en el pretest (Tabla 14-3; Figura 14-1B). De los 2 606 estudiantes matriculados, 184 no completaron el postest. La pérdida no dependió de la condición (OR = 0.90, IC 95 % [0.69, 1.18], p = .448) y fue mayor en el ámbito rural (OR de completar = 0.55, IC 95 % [0.41, 0.75], p < .001); el nivel inicial de lectura crítica tuvo una asociación marginal con completar (OR = 0.87 por desviación estándar, p = .051). Los modelos controlan el ámbito rural y el pretest. La Figura 14-1A resume el flujo y la Figura 14-1C muestra que las dos aulas de cada institución partían de niveles de lectura crítica parecidos.

Tabla 14-3

Características de los participantes y equivalencia inicial entre condiciones (N = 2 422)

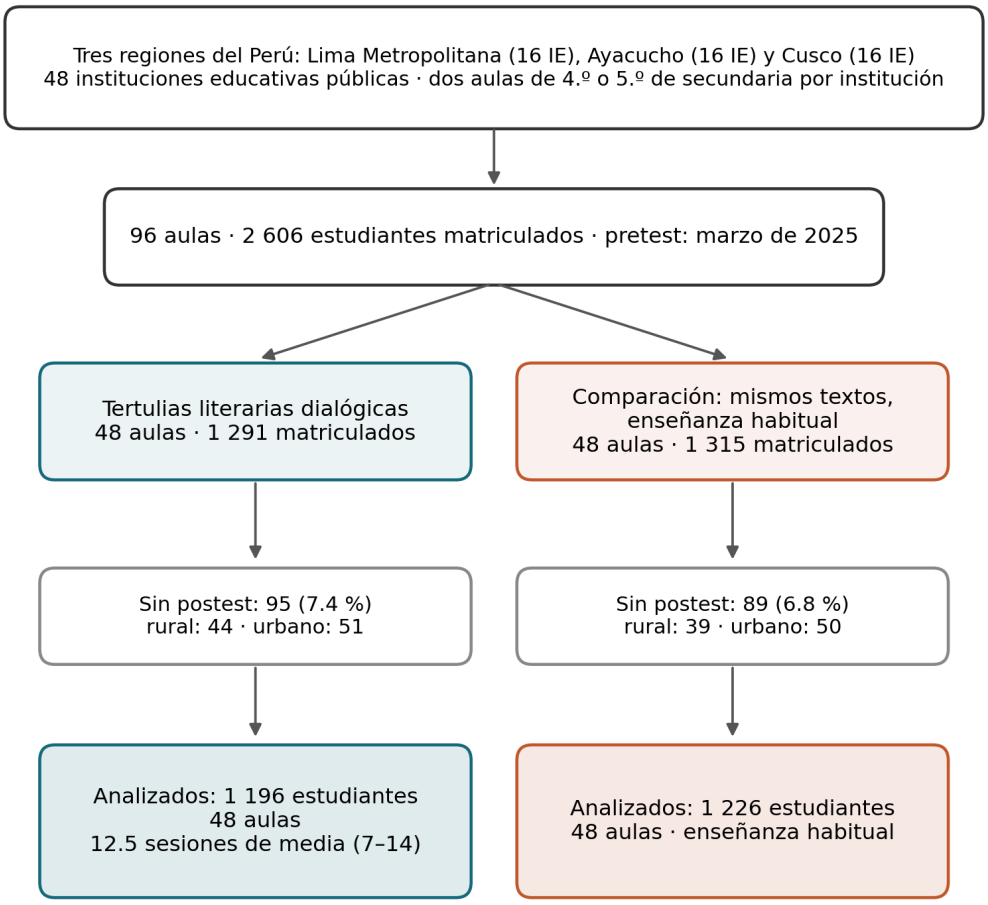
Característica	Tertulias dialógicas	Comparación	Diferencia estandarizada
	(n = 1 196; 48 aulas)		
Mujeres (%)	49.4	51.2	–0.036
Lengua materna quechua (%)	37.5	38.4	–0.020
IE en ámbito rural (%)	32.6	31.8	0.017
Quinto grado (%)	55.2	49.8	0.107
Lectura crítica, pretest (0–10)	4.53 (1.52)	4.69 (1.49)	–0.108
Empatía literaria intercultural, pretest (1–5)	3.07 (0.83)	3.19 (0.85)	–0.152
Pérdida entre oleadas (%)	7.4	6.8	0.023

Nota: Para las variables continuas se informa media (desviación estándar). La diferencia estandarizada es la diferencia de medias (o de proporciones) dividida por la desviación estándar combinada; los valores absolutos menores de 0.20 indican un desequilibrio pequeño. La pérdida entre oleadas se calcula sobre los 2 606 estudiantes matriculados en las 96 aulas.

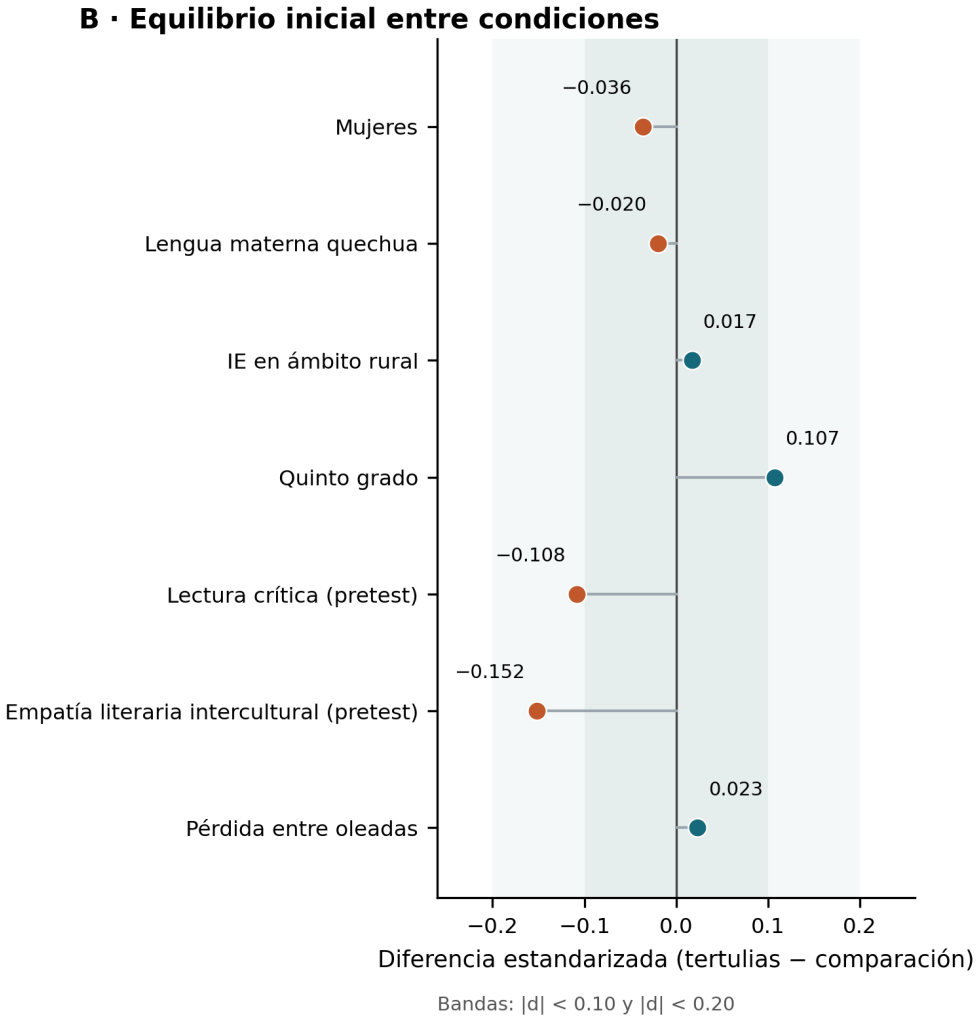
Figura 14-1

Diseño, flujo de participantes y equivalencia inicial de las condiciones

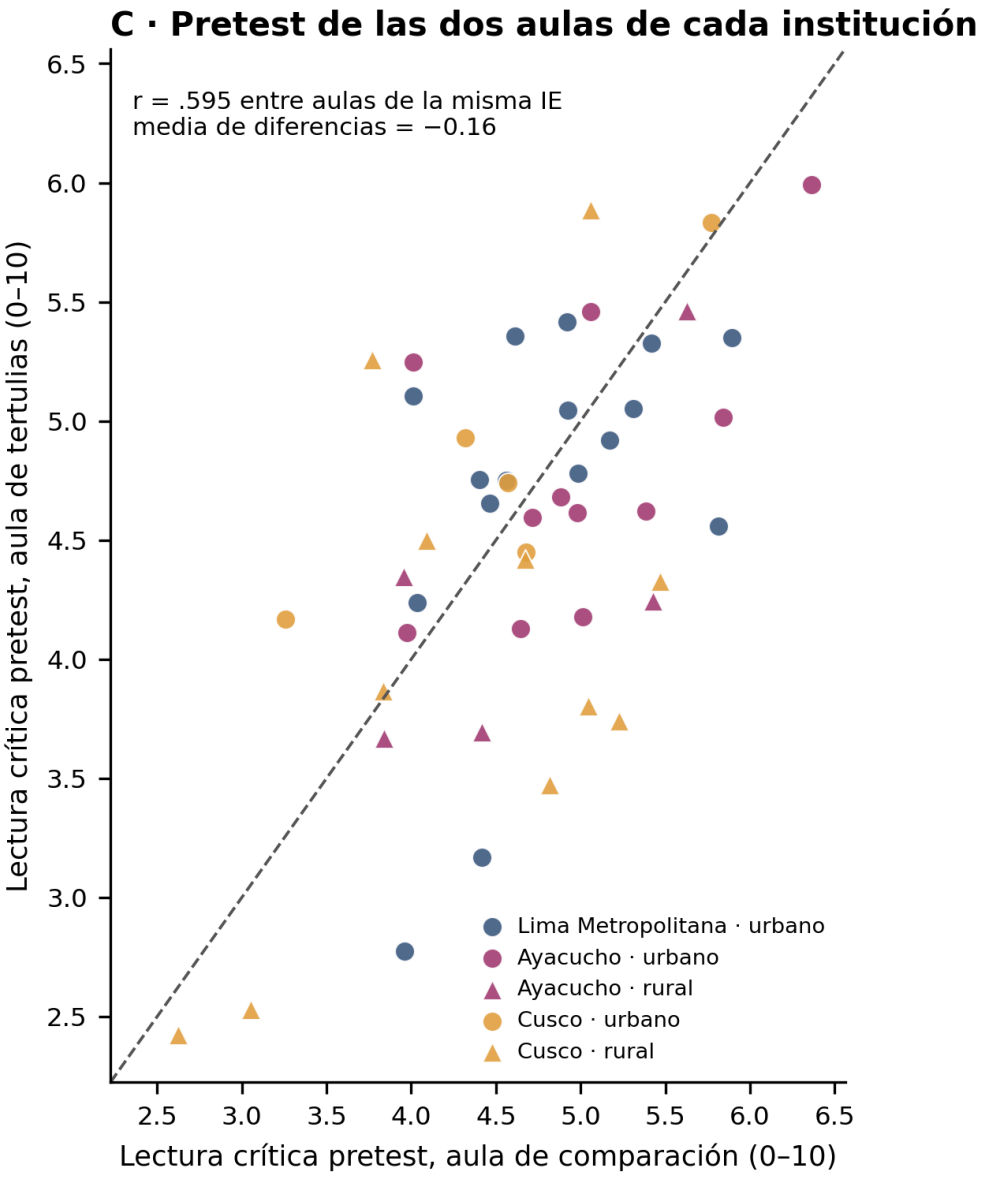
A · Flujo de instituciones, aulas y estudiantes



Nota. Cuasiexperimento pretest-postest con asignación por aula; N = 2 422 estudiantes analizados en 96 aulas de 48 instituciones. Panel A: flujo desde las 48 instituciones educativas (IE) de Lima Metropolitana, Ayacucho y Cusco y sus 96 aulas hasta los 2 422 estudiantes analizados (1 196 con tertulias y 1 226 en comparación); 184 estudiantes no completaron el postest.



Nota. Cuasiexperimento pretest-postest con asignación por aula; N = 2 422 estudiantes analizados en 96 aulas de 48 instituciones. Panel B: diferencias estandarizadas en el pretest entre las condiciones (tertulias – comparación; Tabla 14-3); la mayor en valor absoluto fue –0.152.



Nota. Cuasiexperimento pretest-postest con asignación por aula; N = 2 422 estudiantes analizados en 96 aulas de 48 instituciones. Panel C: media de lectura crítica en el pretest de las dos aulas de cada institución, por región y ámbito; la línea discontinua indica igualdad (r = .595).

14.4.2 Medición: fiabilidad, validez e invarianza

Las tareas y los ítems tuvieron medias centrales y dispersión suficiente, y la concordancia entre evaluadores fue alta en las doce tareas (Tabla 14-2). El modelo de medida de cuatro factores ajustó bien (Tabla 14-4): las cargas estandarizadas oscilaron entre .685 y .787, la fiabilidad compuesta entre .850 y .891 y la varianza media extraída superó .50 en los cuatro factores; las cargas fueron similares en las dos oleadas (Figura 14-2A).

Tabla 14-4

Modelo de medida: cargas factoriales estandarizadas, fiabilidad compuesta y varianza media extraída por oleada (N = 2 422)

Constructo	Indicador	λ pretest	λ postest
Lectura crítica	T1	.764	.787
	T2	.723	.734
	T3	.707	.733
	T4	.742	.779
	T5	.729	.737
	T6	.777	.781
Empatía literaria intercultural	E1	.741	.709
	E2	.760	.751
	E3	.698	.685
	E4	.757	.777
	E5	.701	.721
Lectura crítica	ω (fiabilidad compuesta)	.879	.891
	Varianza media extraída	.548	.576
	α de Cronbach	.879	.890
Empatía literaria intercultural	ω (fiabilidad compuesta)	.852	.850
	Varianza media extraída	.536	.532
	α de Cronbach	.852	.849

Nota: Análisis factorial confirmatorio de cuatro factores (lectura crítica y empatía en cada oleada) con cargas libres y covarianzas entre los residuos del mismo indicador en las dos oleadas; estimación por máxima verosimilitud. Ajuste: $\chi^2(192) = 549.3$, CFI = .986, TLI = .983, RMSEA = .028. T = tarea de la prueba de lectura crítica; E = ítem de la escala de empatía.

La invarianza escalar se sostuvo en los dos constructos (Tabla 14-5; Figura 2B): ninguna restricción empeoró el ajuste de forma significativa ($p \geq .429$) y el CFI cambió como máximo 0.0005.

Tabla 14-5

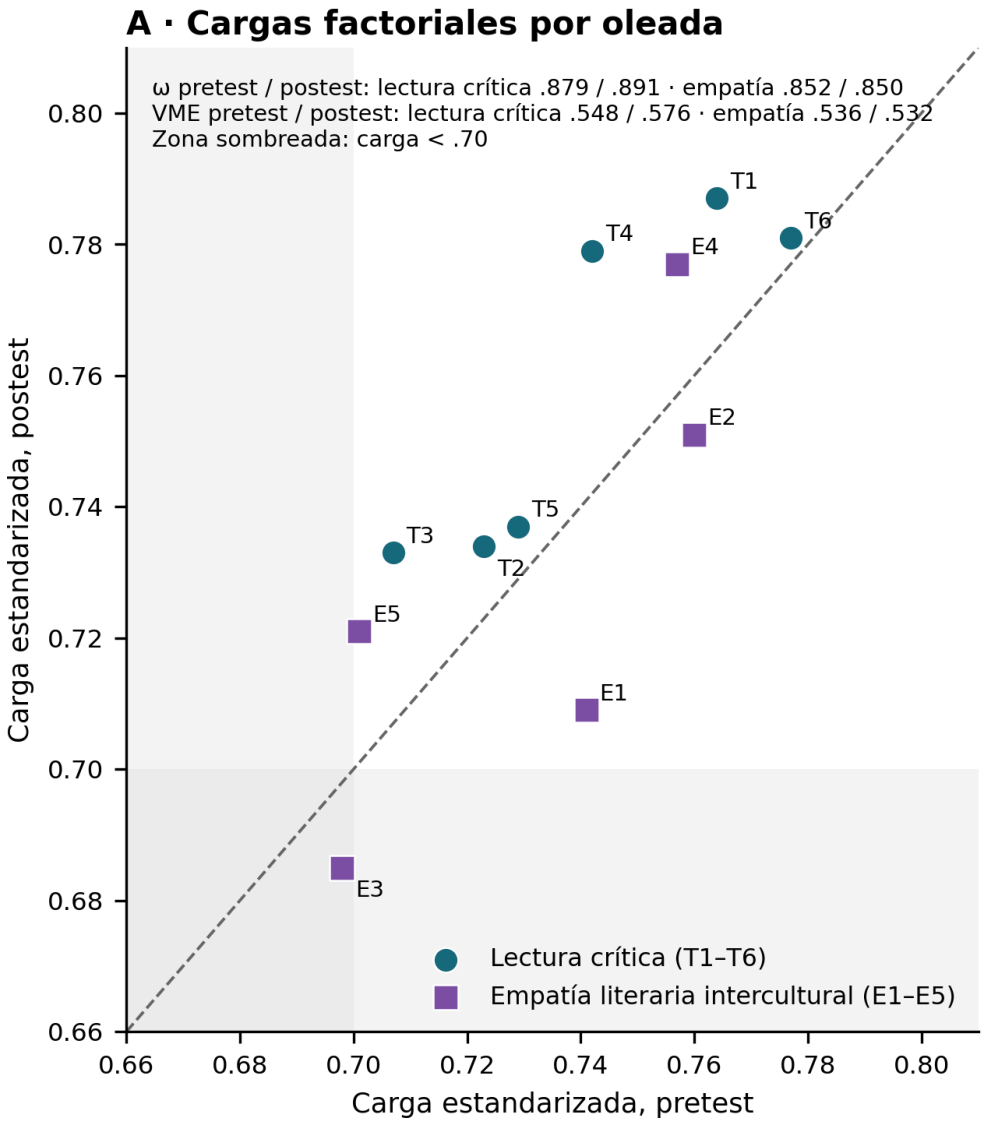
Invarianza de medida longitudinal (pretest-postest) y entre condiciones (tertulias dialógicas frente a comparación)

Constructo	Modelo	χ^2	gl	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ (Δgl)	p	ΔCFI
Lectura crítica	Configural	277.50	94	.988	.983	.040	.022	—	—	—
	Métrica (cargas iguales)	290.10	109	.988	.985	.037	.027	12.59 (15)	.634	0.0002
	Escalar (cargas e interceptos iguales)	298.28	124	.988	.988	.034	.027	8.18 (15)	.916	0.0005
Empatía literaria intercultural	Configural	200.31	58	.986	.979	.045	.021	—	—	—
	Métrica (cargas iguales)	210.27	70	.986	.983	.041	.025	9.96 (12)	.619	0.0002
	Escalar (cargas e interceptos iguales)	222.48	82	.986	.985	.038	.025	12.21 (12)	.429	0.0000

Nota: Modelos de dos grupos con estructura de medias estimados por máxima verosimilitud; cada uno combina la restricción entre oleadas y entre condiciones. Δ = cambio respecto del modelo anterior. El RMSEA de los modelos de dos grupos se multiplica por $\sqrt{2}$ (número de grupos).

Figura 14-2

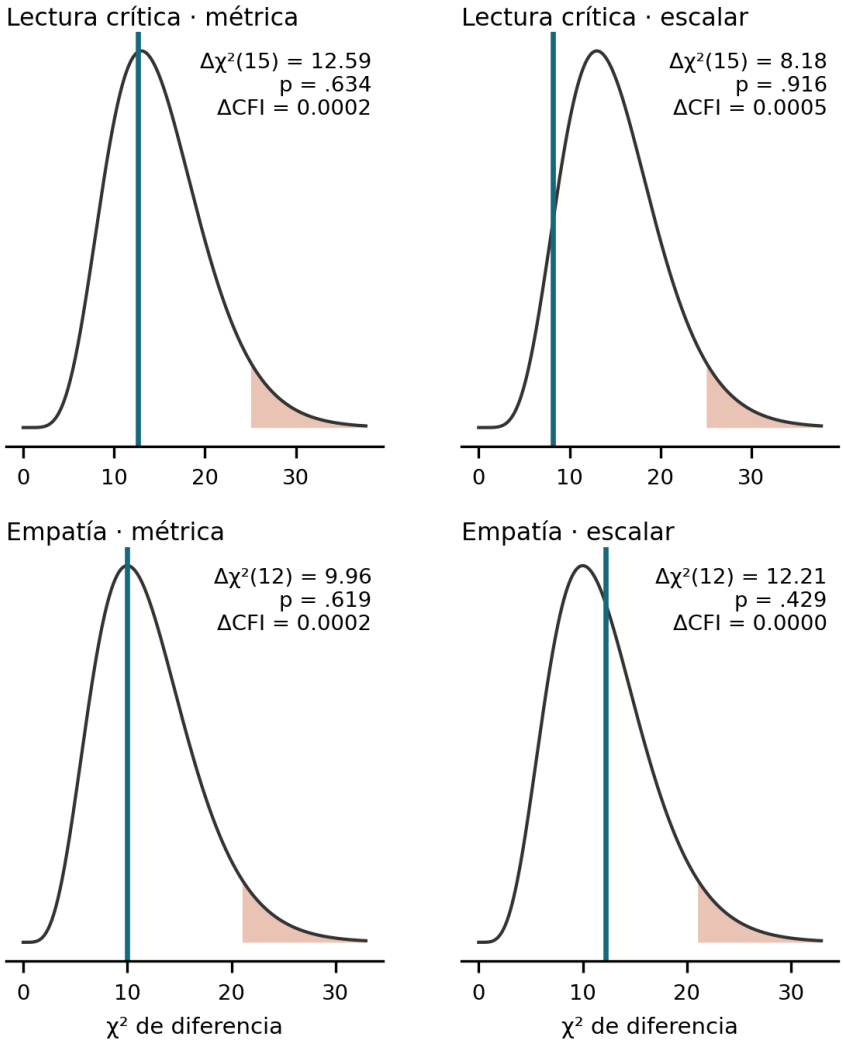
Fiabilidad, invarianza de medida y validez discriminante de las escalas.



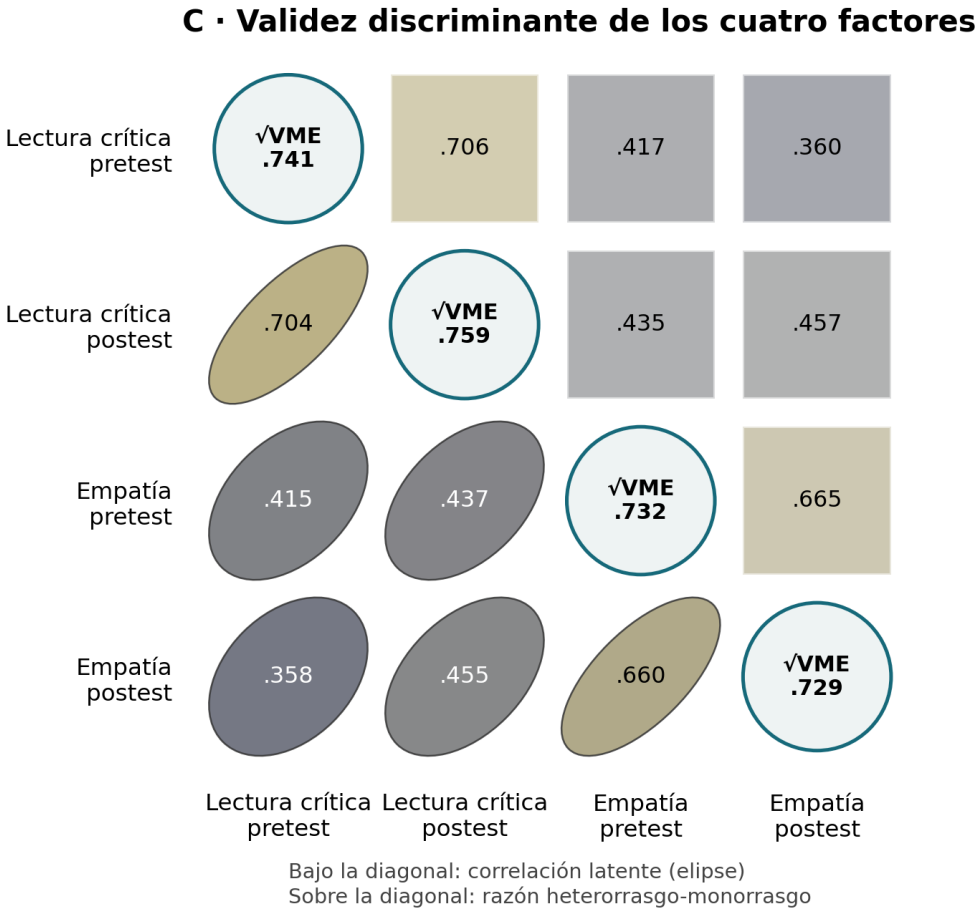
Nota. Análisis factorial confirmatorio por máxima verosimilitud y contrastes de invarianza de dos grupos; N = 2 422 estudiantes. Panel A: cargas estandarizadas del pretest frente al postest (Tabla 14-4) de las seis tareas de lectura crítica (T1–T6) y los cinco ítems de empatía literaria intercultural (E1–E5); la diagonal indica igualdad. ω = fiabilidad compuesta (lectura crítica .879 y .891; empatía .852 y .850); VME = varianza media extraída.

B · Contrastes de invarianza frente a su distribución χ^2 de referencia

Región sombreada: 5 % superior · línea: $\Delta\chi^2$ observado



Nota. Análisis factorial confirmatorio por máxima verosimilitud y contrastes de invarianza de dos grupos; N = 2 422 estudiantes. Panel B: diferencia de χ^2 de los contrastes de invarianza métrica y escalar (Tabla 5) sobre su distribución χ^2 de referencia; todos con $p \geq .429$.



Nota. Análisis factorial confirmatorio por máxima verosimilitud y contrastes de invarianza de dos grupos; N = 2 422 estudiantes. Panel C: raíz cuadrada de la VME en la diagonal, correlaciones latentes bajo ella (la elipse se estrecha al aumentar la correlación) y razón heterorrasgo-monorrasgo sobre ella (Tabla 14-6).

La raíz de la varianza media extraída de cada factor superó todas sus correlaciones latentes, incluida la que une la lectura crítica del pretest y la del postest, y la razón heterorrasgo-monorrasgo máxima fue .706 (Tabla 14-6; Figura 14-2C).

Tabla 14-6

Validez discriminante

Constructo			1	2	3	4
1.	Lectura crítica	pretest	**.741**	.706	.417	.360
2.		postest	.704	**.759**	.435	.457
3.	Empatía	pretest	.415	.437	**.732**	.665
4.		postest	.358	.455	.660	**.729**

Nota: Raíz cuadrada de la varianza media extraída (diagonal), correlaciones latentes (bajo la diagonal) y razón heterorrasgo-monorrasgo (sobre la diagonal). Correlaciones latentes del análisis factorial confirmatorio de la Tabla 4. La razón heterorrasgo-monorrasgo se calcula con las correlaciones entre indicadores (Henseler et al., 2015).

14.4.3 Descriptivos y diferencias ajustadas

La lectura crítica aumentó 0.50 puntos en las aulas de tertulias y 0.11 en las de comparación; la diferencia ajustada se repitió en las tres dimensiones, con d entre 0.17 y 0.20, y la empatía aumentó 0.20 puntos con las tertulias y 0.02 en la comparación (Tabla 14-7). La correlación intraclase de aula de la lectura crítica fue .246 en el pretest y .290 en el postest.

Tabla 14-7

Descriptivos Descriptivos por condición y oleada, y diferencia ajustada en el postest (N = 2 422)

Variable	TLD pretest M (DE)		Comparación M (DE)		Diferencia ajustada B [IC 95 %]	d	CCI aula (postest)
	pretest	postest	pretest	postest			
Lectura crítica, puntuación global (0–10)	4.53 (1.52)	5.03 (1.58)	4.69 (1.49)	4.80 (1.58)	0.34 [0.20, 0.47]	0.22	.290
Interpretación inferencial (T1–T2)	5.28 (1.70)	5.77 (1.78)	5.46 (1.65)	5.54 (1.73)	0.33 [0.18, 0.48]	0.20	.241
Análisis ideológico-intercultural (T3–T4)	4.39 (1.68)	4.92 (1.75)	4.54 (1.65)	4.66 (1.76)	0.34 [0.18, 0.50]	0.20	.248
Evaluación argumentada (T5–T6)	3.90 (1.74)	4.40 (1.77)	4.06 (1.72)	4.18 (1.80)	0.29 [0.11, 0.47]	0.17	.216
Empatía literaria intercultural (1–5)	3.07 (0.83)	3.26 (0.83)	3.19 (0.85)	3.21 (0.86)	0.12 [0.07, 0.18]	0.15	.121

Nota: TLD = tertulias literarias dialógicas. Diferencia ajustada por el pretest de la misma variable, el ámbito rural y la lengua materna, con errores estándar robustos por institución educativa (48 conglomerados); d = diferencia ajustada dividida por la desviación estándar del pretest. Las tres dimensiones son promedios de pares de tareas de la puntuación global. CCI = correlación intraclase de aula del postest.

14.4.4 Modelo estructural

El modelo estructural latente ajustó bien: $\chi^2(316) = 679.0$, CFI = .986, TLI = .984, RMSEA = .022 (IC 90 % [.020, .024]) y SRMR = .022. Las tertulias aumentaron la empatía postest controlando su pretest y la lectura crítica inicial (B = 0.130, p < .001; d = 0.17). La empatía postest predijo la lectura crítica postest (β = .190, p < .001) y la condición mantuvo un efecto directo (B = 0.332, p < .001; d = 0.21). El modelo explicó el 45.9 % de la varianza de la empatía postest y el 56.0 % de la lectura crítica postest (Tabla 14-8; Figura 14-3A).

Tabla 14-8

Modelo de ecuaciones estructurales latente de dos oleadas: coeficientes estructurales (N = 2 422)

Variable dependiente	Predictor	B	EE	p	β	R^2
Empatía literaria intercultural (postest)	Tertulias dialógicas (1 = sí)	0.130	0.028	< .001	.083 (d = 0.17)	.459
	Empatía pretest	0.623	0.022	< .001	.616	
	Lectura crítica pretest	0.057	0.012	< .001	.107	
	Lengua materna quechua	–0.005	0.029	.867	–.003	
	Mujer	0.080	0.028	.004	.052	
Lectura crítica (postest)	Tertulias dialógicas (1 = sí)	0.332	0.050	< .001	.107 (d = 0.21)	.560
	Empatía postest	0.379	0.052	< .001	.190	
	Lectura crítica pretest	0.644	0.021	< .001	.603	
	Empatía pretest	0.129	0.053	.016	.064	
	Ámbito rural	–0.164	0.054	.002	–.049	
Lectura crítica (pretest)	Lengua materna quechua	–0.145	0.065	.025	–.049	.024
	Ámbito rural	–0.429	0.064	< .001	–.138	
	Quinto grado	0.003	0.058	.958	.001	
Empatía (pretest)	Mujer	0.119	0.032	< .001	.077	.007
	Lengua materna quechua	–0.057	0.035	.101	–.036	

Nota: Estimación por máxima verosimilitud con cargas iguales entre oleadas (invarianza métrica) y covarianzas residuales del mismo indicador. B = coeficiente no estandarizado; β = coeficiente estandarizado; para la condición se informa además d, el coeficiente expresado en desviaciones estándar de la variable dependiente. Ajuste global: $\chi^2(316) = 679.0$, $p < .001$; CFI = .986; TLI = .984; RMSEA = .022, IC 90 % [.020, .024]; SRMR = .022.

14.4.5 Efectos total, directo e indirecto

El efecto total de las tertulias sobre la lectura crítica latente fue d = 0.25 (IC 95 % bootstrap [0.16, 0.33]), frente a d = 0.22 en el compuesto observado (Tabla 14-7). El intervalo del efecto indirecto a través de la empatía excluyó el cero; este representó el 12.9 % del efecto total (IC 95 % [7.6, 20.8]), y el efecto directo siguió siendo d = 0.21 (IC 95 % [0.13, 0.30]) (Tabla 14-9). Los 1 000 remuestreos convergieron; las distribuciones bootstrap y

la comparación con la aproximación de Monte Carlo se muestran en las Figuras 14-3B y 14-3C.

Tabla 14-9

Descomposición del efecto de las tertulias dialógicas sobre la lectura crítica: efectos total, directo e indirecto a través de la empatía literaria intercultural

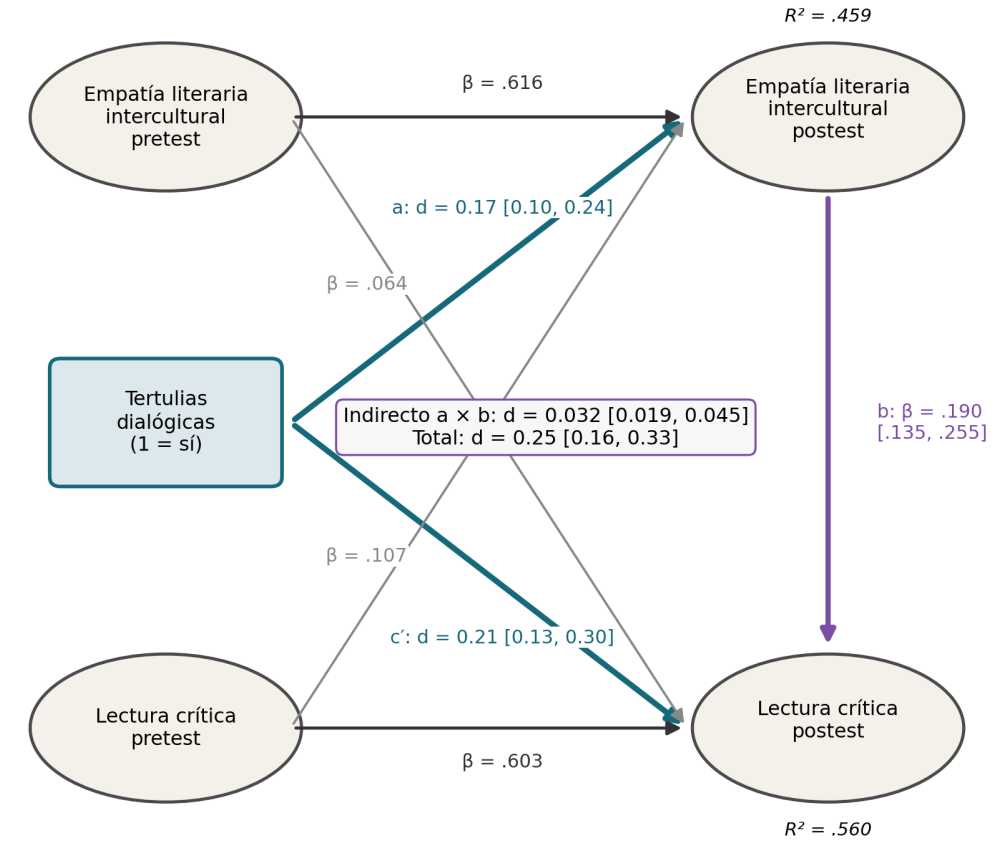
Efecto	Estimación (B)	IC 95 % bootstrap por institución	Estimación estandarizada	IC 95 % bootstrap (estandarizada)
Ruta a: tertulias → empatía (postest)	0.130	[0.074, 0.187]	0.166	[0.096, 0.237]
Ruta b: empatía (postest) → lectura crítica (postest)	0.379	[0.261, 0.514]	0.190	[0.135, 0.255]
Efecto directo c′: tertulias → lectura crítica	0.332	[0.195, 0.458]	0.215	[0.127, 0.302]
Efecto indirecto a × b	0.049	[0.029, 0.071]	0.032	[0.019, 0.045]
Efecto total c = c′ + a × b	0.381	[0.245, 0.509]	0.246	[0.157, 0.333]
Proporción mediada (a × b / c)	.129	[.076, .208]	—	—

Nota: Intervalos percentiles de 1 000 remuestreos con reemplazo de las 48 instituciones educativas, que conservan juntas las dos aulas de cada institución. Las estimaciones estandarizadas de a, c′, a × b y c están en desviaciones estándar de la variable dependiente (d); la de b es un β. El intervalo de Monte Carlo del efecto indirecto basado en los errores estándar del modelo es [0.026, 0.075].

Figura 14-3

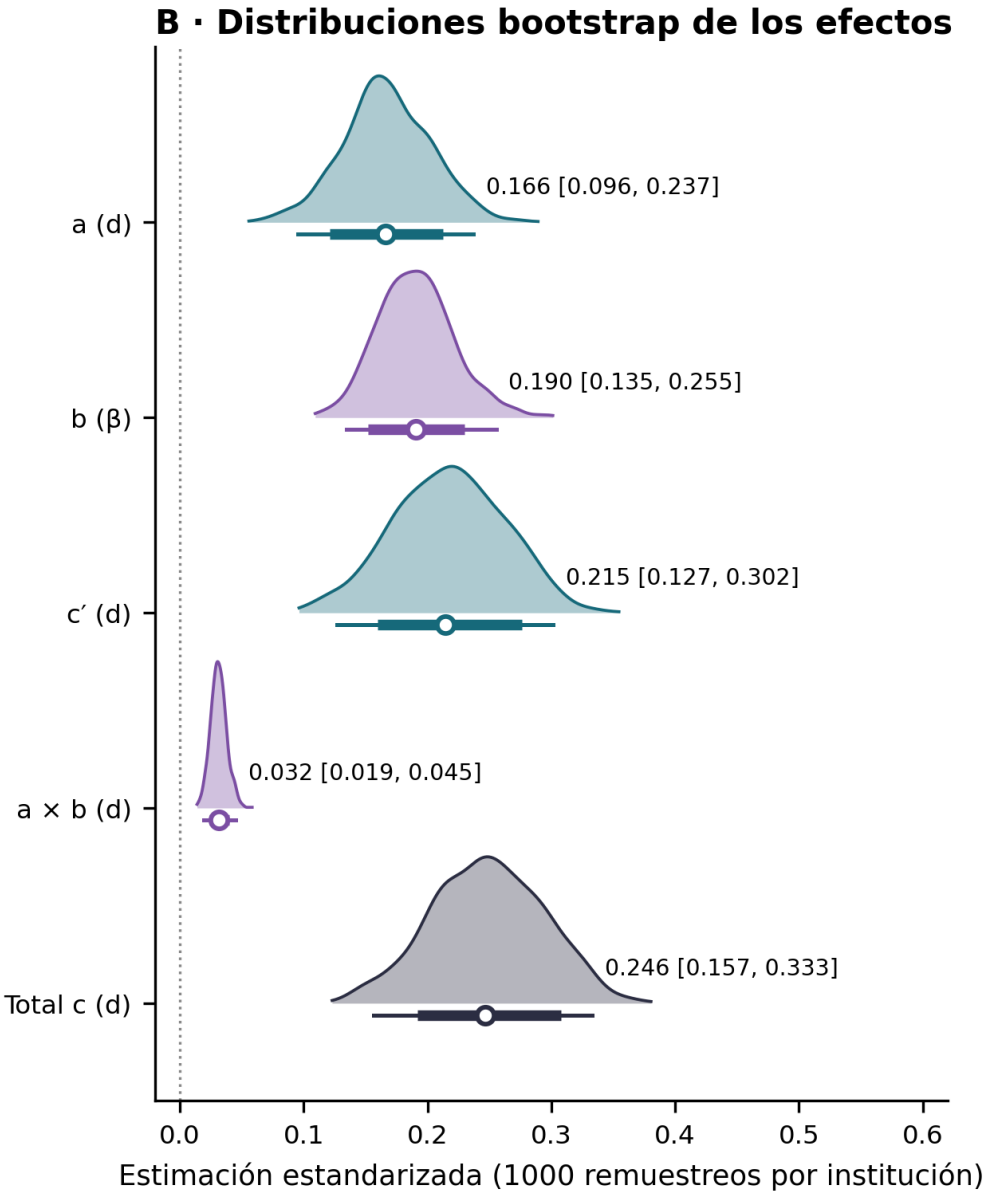
Modelo estructural latente y descomposición del efecto de las tertulias sobre la lectura crítica

A · Modelo estructural latente de dos oleadas

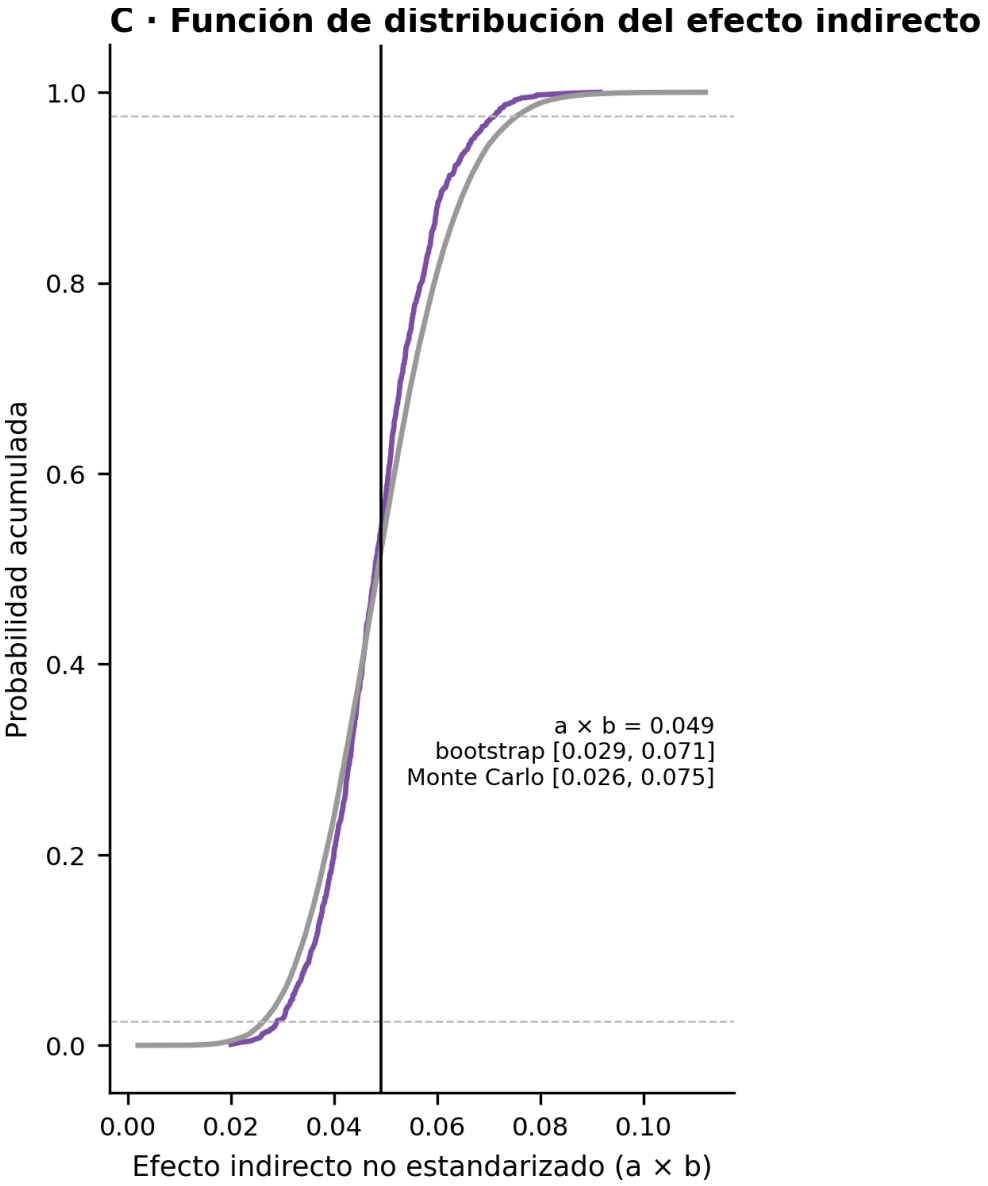


χ²(316) = 679.0 · CFI = .986 · TLI = .984 · RMSEA = .022 [.020, .024] · SRMR = .022 · covariables: lengua materna, sexo, ámbito rural, grado

Nota. Modelo de ecuaciones estructurales latente de dos oleadas por máxima verosimilitud, con remuestreo bootstrap de instituciones; N = 2 422 estudiantes en 48 instituciones. Panel A: estimaciones del modelo de dos oleadas (Tabla 14-8 y Tabla 14-9); d en desviaciones estándar de la variable dependiente, β estandarizado e intervalos bootstrap del 95 % por institución; $\chi^2(316) = 679.0$, CFI = .986, RMSEA = .022.



Nota. Modelo de ecuaciones estructurales latente de dos oleadas por máxima verosimilitud, con remuestreo bootstrap de instituciones; N = 2 422 estudiantes en 48 instituciones. Panel B: densidades de 1 000 remuestreos de las 48 instituciones; trazo grueso, intervalo del 80 %; trazo fino, intervalo del 95 %; círculo, estimación puntual.



Nota. Modelo de ecuaciones estructurales latente de dos oleadas por máxima verosimilitud, con remuestreo bootstrap de instituciones; N = 2 422 estudiantes en 48 instituciones. Panel C: función de distribución acumulada del efecto indirecto no estandarizado (a × b = 0.049) en el bootstrap (IC 95 % [0.029, 0.071]) y en la aproximación de Monte Carlo (IC 95 % [0.026, 0.075]); las líneas horizontales marcan los percentiles 2.5 y 97.5.

14.4.6 Robustez y heterogeneidad

El efecto sobre la lectura crítica se mantuvo en todas las especificaciones alternativas, con estimaciones entre d = 0.23 y d = 0.26 (Tabla 14-10; Figura 14-4B). Al excluir sucesivamente cada región, el efecto total latente osciló entre 0.21 y 0.27 y el indirecto entre 0.029 y 0.034.

Tabla 14-10

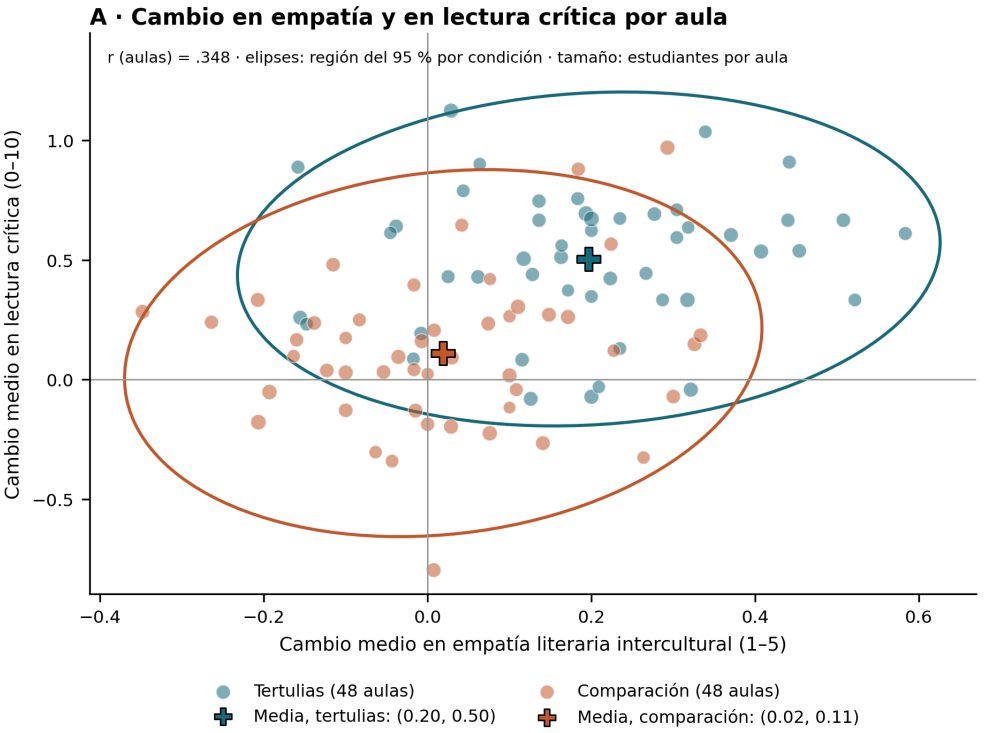
Robustez: especificaciones alternativas)

Especificación		Estimación (d)	IC 95 %	p
Modelo multinivel de tres niveles (estudiantes, aulas e instituciones), puntuación global		0.233	[0.153, 0.314]	< .001
Puntuación de ganancia (diferencia en diferencias)		0.260	[0.181, 0.340]	< .001
Ponderación por la probabilidad inversa de la condición		0.247	[0.158, 0.336]	< .001
Efectos fijos de institución (comparación dentro de cada institución)		0.233	[0.137, 0.329]	< .001
Modelo estructural sin la región de	Lima Metropolitana: efecto total (indirecto = 0.034)	0.209	—	< .001
	Ayacucho: efecto total (indirecto = 0.029)	0.266	—	< .001
	Cusco: efecto total (indirecto = 0.032)	0.262	—	< .001
Interacción tratamiento	× lengua materna quechua sobre lectura crítica	–0.049	[–0.203, 0.106]	.537
	× lengua materna quechua sobre empatía	0.007	[–0.130, 0.144]	.919
	× ámbito rural sobre lectura crítica	–0.117	[–0.296, 0.061]	.197
Interacción tratamiento × ámbito rural sobre empatía		0.008	[–0.137, 0.154]	.910
Dosis: sesiones asistidas en el grupo de tertulias (por sesión)		0.003	[–0.038, 0.045]	.876

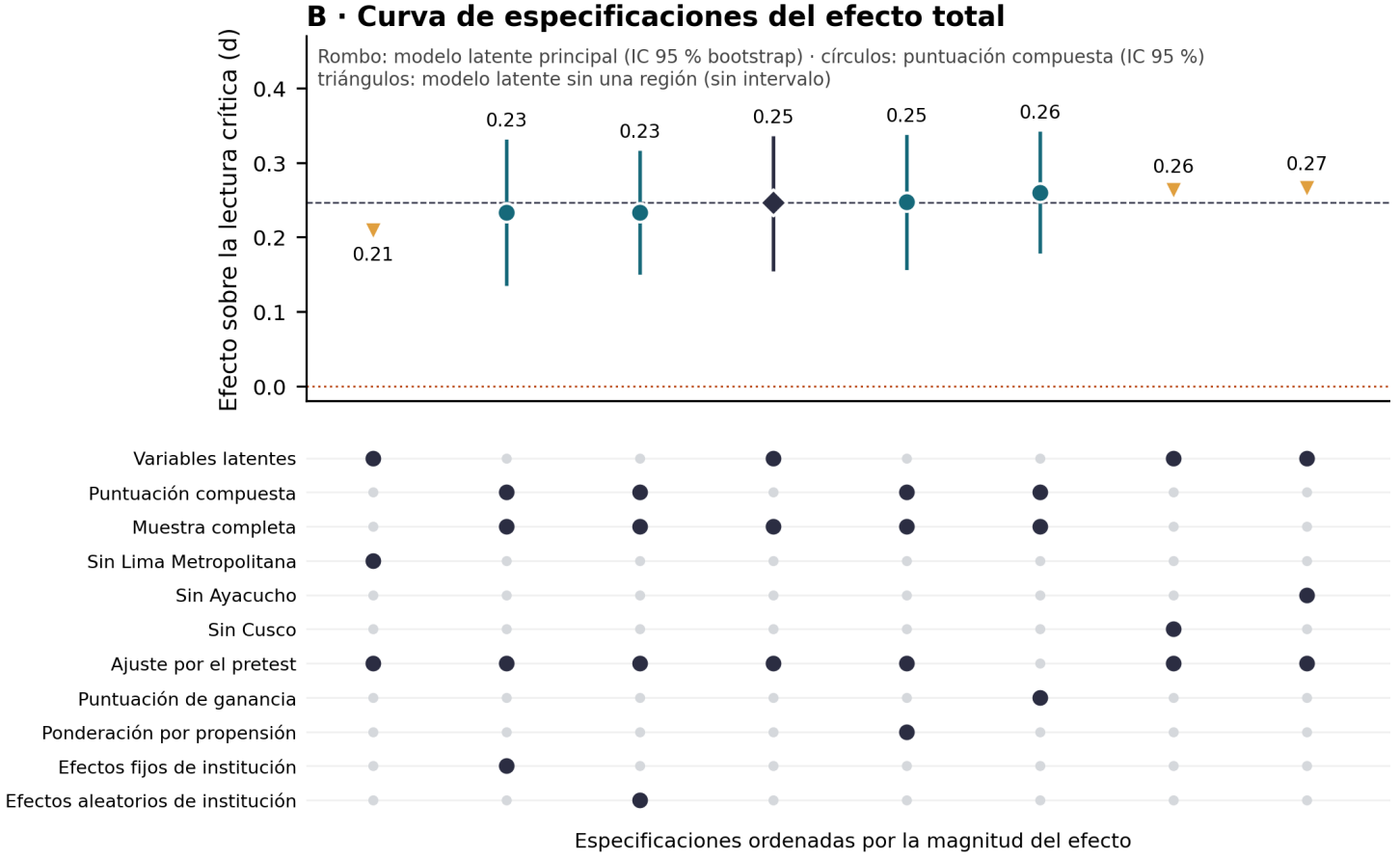
Nota: Todas las estimaciones se expresan en desviaciones estándar del pretest de la variable de resultado. Los modelos de regresión usan errores estándar robustos por institución educativa; el multinivel incluye interceptos aleatorios de institución y de aula; la ponderación usa una puntuación de propensión estimada con los pretests y las covariables (rango .356–.650). En las filas de exclusión de una región, p corresponde al efecto directo.

Figura 14-4

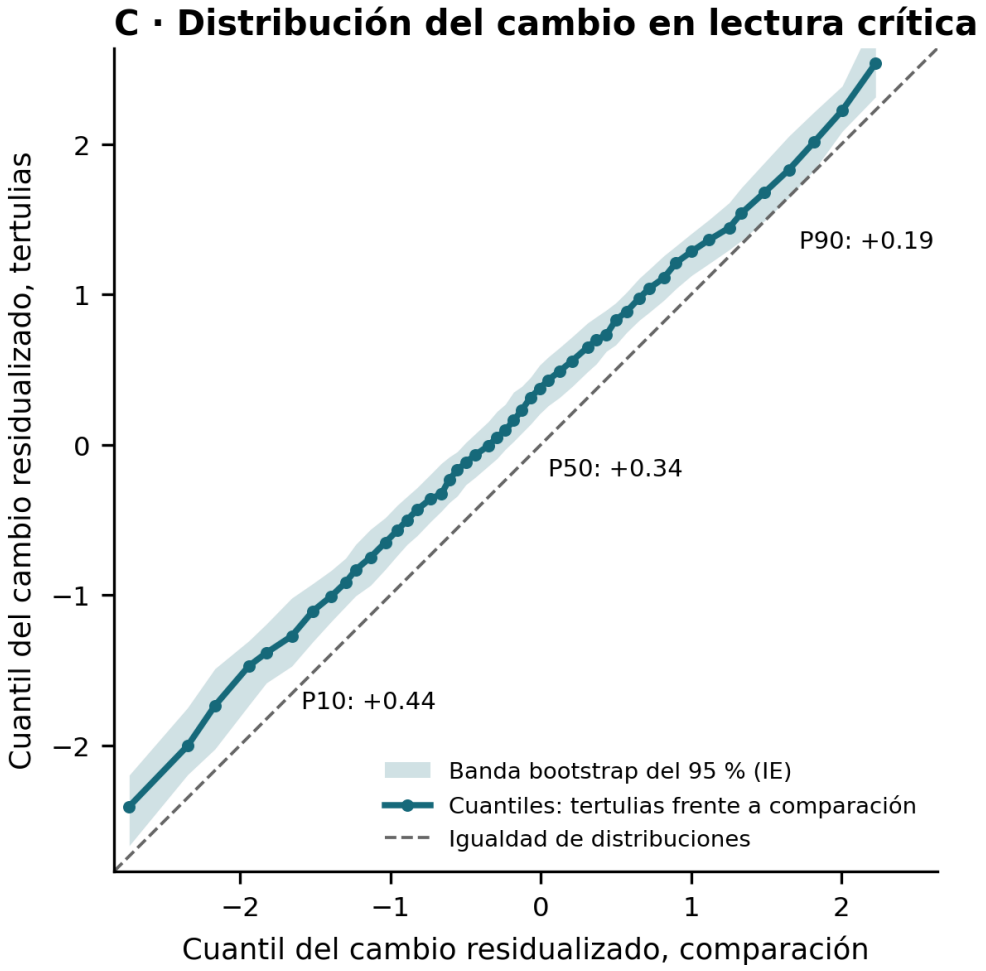
Robustez y heterogeneidad del efecto de las tertulias



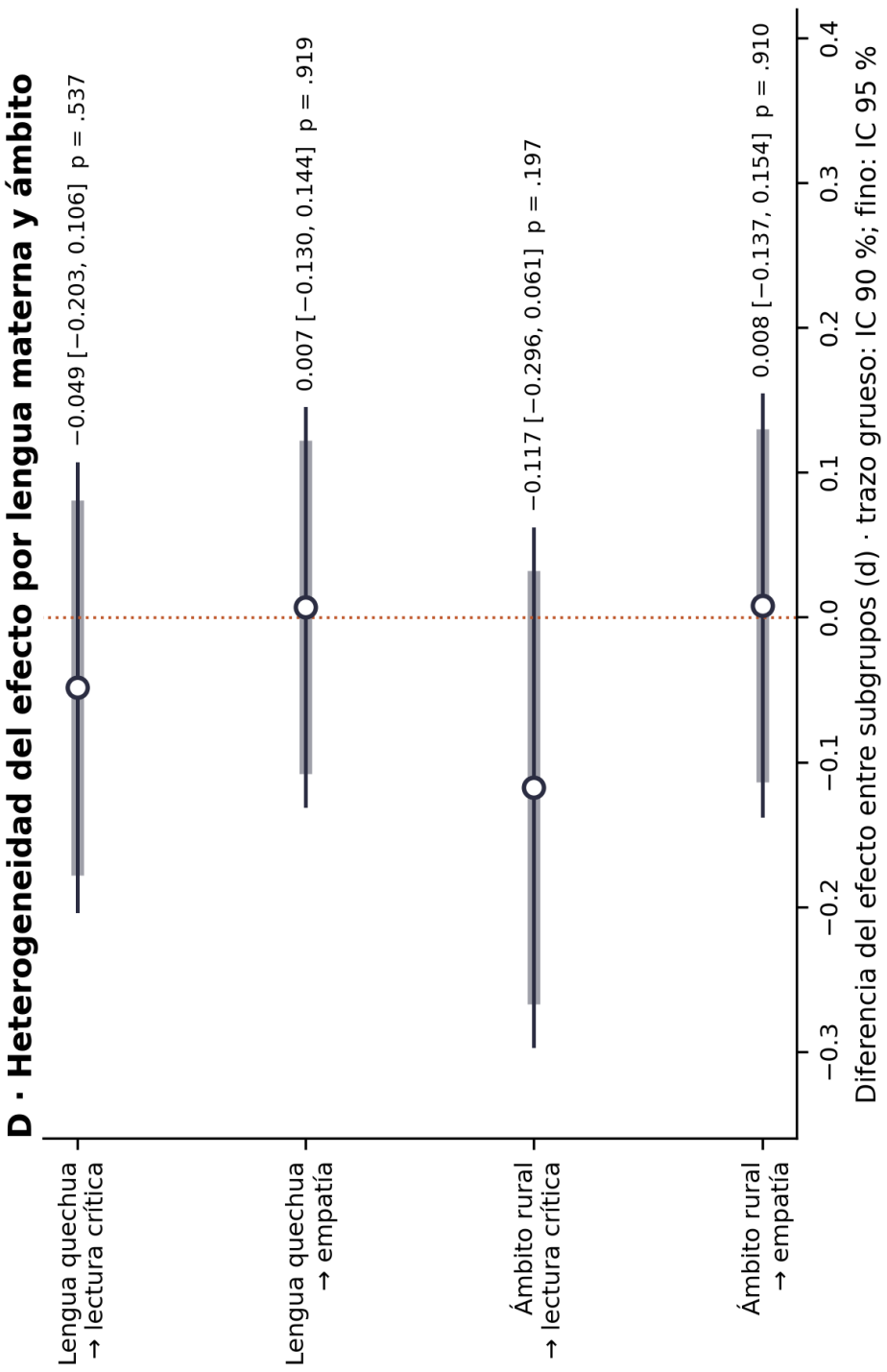
Nota. Especificaciones de robustez y análisis de heterogeneidad del efecto; N = 2 422 estudiantes en 96 aulas de 48 instituciones. Panel A: cambio medio de pretest a postest en empatía literaria intercultural y en lectura crítica en cada una de las 96 aulas; las cruces marcan la media de cada condición y las elipses, la región del 95 % (r = .348 entre aulas).



Nota. Especificaciones de robustez y análisis de heterogeneidad del efecto; N = 2 422 estudiantes en 96 aulas de 48 instituciones. Panel B: efecto total sobre la lectura crítica en las 8 especificaciones (Tabla 14-9 y Tabla 14-10), ordenadas de menor a mayor (d entre 0.21 y 0.27); la matriz inferior indica los rasgos de cada especificación.



Nota. Especificaciones de robustez y análisis de heterogeneidad del efecto; N = 2 422 estudiantes en 96 aulas de 48 instituciones. Panel C: gráfico cuantil-cuantil del cambio en lectura crítica residualizado respecto del pretest; desplazamientos de 0.44, 0.34 y 0.19 puntos en los percentiles 10, 50 y 90, con banda bootstrap del 95 % por institución.



Nota. Especificaciones de robustez y análisis de heterogeneidad del efecto; N = 2 422 estudiantes en 96 aulas de 48 instituciones. Panel D: diferencia del efecto entre subgrupos de lengua materna y de ámbito (Tabla 14-10); trazo grueso, IC 90 %; trazo fino, IC 95 %; todas con p ≥ .197..

Ninguna interacción con la lengua materna quechua ni con el ámbito rural fue significativa ($p \geq .197$). Dentro del grupo de tertulias, el número de sesiones asistidas no se asoció con la lectura crítica final. Por aula, el cambio en empatía y el cambio en lectura crítica fueron mayores con las tertulias y estuvieron positivamente relacionados (Figura 14-4A); la ventaja se observó en toda la distribución del cambio residualizado, algo más amplia en la parte baja (Figura 14-4C), y las diferencias entre subgrupos se muestran en la Figura 14-4D.

14.5 DISCUSIÓN

El hallazgo más desconcertante no es que las tertulias mejoraran la lectura crítica, sino lo poco que de esa mejora viaja por la ruta que la teoría señalaba como central: la empatía literaria intercultural canalizó en torno a una octava parte del efecto total, y la mayor parte permaneció como efecto directo. A ello se suma una segunda anomalía, la ausencia de gradiente con la asistencia, dentro de un rango estrecho de sesiones. Ambos resultados, tomados juntos, sugieren que el beneficio no depende de acumular sesiones de forma lineal ni se explica sobre todo por la comprensión de los personajes. Una lectura plausible, aunque no contrastada aquí, es que la propia estructura de la tertulia —elegir un pasaje, defenderlo y ser contradicho con argumentos— entrena desde las primeras sesiones la práctica argumentativa que la prueba mide en sus tres dimensiones, y que la empatía añade una contribución real pero secundaria. Esta interpretación es consistente con la evidencia de que la argumentación literaria dialógica favorece la escritura argumentativa y las creencias motivacionales (Fulton et al., 2023) y con el papel mediador de la motivación y la autoeficacia en la comprensión lectora (Moulavinafchi et al., 2026), vías que este estudio no midió.

En contraste con las síntesis disponibles, la magnitud del efecto total ($d = 0.25$) queda por debajo de los efectos grandes que informan los metaanálisis de lectura dialógica en la infancia (Tang et al., 2026) y de círculos literarios en lectores de segunda lengua (Ma, 2025), y por encima del efecto pequeño de la formación docente en comprensión lectora (Rice

et al., 2024). Esa posición intermedia no debe leerse como una debilidad de la actuación, sino como una consecuencia del diseño: la condición de comparación trabajó los mismos textos en el mismo periodo, de modo que el contraste aísla la forma de discutirlos y no la exposición a la literatura, y la medida fue una prueba de respuesta construida ajena a las sesiones, más exigente que las medidas próximas a la intervención con las que suelen obtenerse los efectos mayores. Que la ventaja apareciera también en el análisis ideológico-intercultural y en la evaluación argumentada, y no solo en la interpretación inferencial, matiza la conclusión de que muy pocas modalidades de discusión mejoran el pensamiento crítico (Murphy et al., 2009). Una posible explicación es que la tertulia, a diferencia de otras formas de conversación sobre el texto, institucionaliza la valoración de las intervenciones por sus argumentos. Esto resuena con la observación de que las posturas lectoras abren o estrechan las interpretaciones según el diálogo que las sostenga (Barak y Lefstein, 2022), aunque este estudio no observó la calidad del diálogo y la conexión es, por tanto, inferencial. Tampoco puede descartarse que parte del efecto responda a diferencias entre docentes no capturadas por la asignación no probabilística, pese a la convergencia de las cinco especificaciones de robustez.

Respecto del mecanismo, la prueba de mediación da contenido empírico a una idea que la literatura sobre tertulias sostenía con evidencia sobre todo cualitativa. El aumento de la empatía (d = 0.17) es del orden de la mejora de la cognición social tras leer ficción (Dodell y Tamir, 2018) y de la toma de perspectiva tras discutir literatura en pequeños grupos (Frei et al., 2026). El resultado también matiza los hallazgos nulos de experimentos preregistrados con lecturas breves (Wimmer et al., 2021): sugiere que no basta con leer sobre otros y que la discusión sostenida, en la que cada voz se escucha por sus argumentos, podría ser lo que convierte la lectura en comprensión de otras perspectivas. La coincidencia, con todo, es parcial, pues los estudios citados miden la cognición social con tareas distintas de un autoinforme sobre personajes andinos.

Desde la perspectiva de la educación intercultural, la ausencia de interacciones con la lengua materna quechua y con el ámbito rural admite

una lectura de equidad, siempre con la cautela de que la ausencia de significación no equivale a equivalencia demostrada: los intervalos de las interacciones son amplios. Aun así, una actuación que no exige recursos costosos produjo un beneficio comparable para estudiantes cuyo mundo representa la literatura indigenista y para el resto, lo que la hace una candidata razonable para el área de Comunicación en un sistema donde las reformas curriculares de la formación docente buscan llevar la diversidad cultural a las aulas pero enfrentan tensiones entre el normativismo funcional y las posturas emergentes (Chávez, 2024; Moreano et al., 2026). La interpretación de fondo es que la literatura que representa a pueblos históricamente subordinados, discutida en diálogo igualitario, puede servir a la vez a la alfabetización crítica y a una educación intercultural que invite a comprender desde dentro, sin subordinar una meta a la otra (Garcia et al., 2023). Comprender por qué actúan como actúan los comuneros frente al gamonalismo exige, además, suspender el estereotipo y reconocer otra lógica, una operación que las propuestas decoloniales describen como inseparable de la pedagogía crítica (Walsh, 2023).

Hipótesis por hipótesis, la secuencia es la siguiente.

1) H1 se sostiene, con un efecto moderado y robusto.

2) H2 se sostiene, con una magnitud algo inferior a la prevista en la planificación.

3) H3 se sostiene, con una asociación significativa pero de tamaño modesto.

4) H4 se sostiene como mediación parcial, con un intervalo que excluye el cero y una proporción mediada pequeña.

La pregunta de equidad se responde con ausencia de evidencia de heterogeneidad. Este patrón sugiere que la teoría del mecanismo empático es válida pero insuficiente, y que una teoría completa de las tertulias debería incorporar la práctica argumentativa y la motivación como vías paralelas.

Conviene reconocer las limitaciones con precisión. Las magnitudes observadas de la ruta a (d = 0.17) y del efecto total (d = 0.25) quedaron algo por debajo de los valores con los que se planificó la muestra (0.20 y 0.30), aunque sus intervalos excluyen el cero; el efecto indirecto se estimó, por tanto, con menos margen del previsto y debe leerse junto a su intervalo. La asignación de las aulas no fue probabilística; la comparación dentro de cada institución, la equivalencia inicial y la convergencia de las especificaciones reducen, pero no eliminan, el riesgo de confusión por diferencias entre docentes. El diseño de dos oleadas establece la secuencia entre condición y mediador, pero la empatía y la lectura crítica finales se midieron a la vez, por lo que la dirección de la ruta b descansa en el control de ambos pretests y en supuestos que deben explicitarse (Kim y Thoemmes, 2025; Valente et al., 2021). La empatía se midió por autoinforme, la pérdida fue mayor en el ámbito rural y las instituciones no proceden de un muestreo probabilístico, de modo que los resultados describen a las instituciones estudiadas.

Las líneas futuras se derivan de esas limitaciones: asignar las aulas por sorteo, añadir una tercera medición del mediador a mitad de la intervención para distinguir mejor su temporalidad, observar la calidad del diálogo en las sesiones, medir la empatía con tareas de desempeño y poner a prueba explícitamente la práctica argumentativa y la motivación como mediadores alternativos. Estudiar rangos de dosis más amplios permitiría, además, aclarar si la falta de gradiente refleja un efecto de umbral o, simplemente, la estrechez del rango observado.

14.6 CONCLUSIONES

Las tertulias literarias dialógicas con narrativa indigenista mejoraron la lectura crítica de estudiantes de secundaria en tres regiones de Perú, con un efecto moderado, robusto a especificaciones alternativas y sin diferencias detectables entre estudiantes de lengua materna quechua o de ámbito rural y el resto. El resultado se apoya en un diseño con comparación activa que trabajó los mismos textos, en medidas latentes con invarianza escalar entre oleadas y condiciones y en una muestra planificada para el contraste

más exigente, lo que fortalece la inferencia sobre el efecto de la forma de discutir la literatura y no solo de leerla.

Una parte modesta pero estimable de ese efecto opera a través de la empatía literaria intercultural, lo que ofrece una prueba de mecanismo que la investigación sobre tertulias no tenía y que conecta la didáctica de la lectura crítica con la educación intercultural. Al mismo tiempo, el predominio del efecto directo indica que existen vías complementarias, probablemente argumentativas y motivacionales, que deberán estudiarse con diseños que midan el mediador en más de un momento y que observen la calidad del diálogo.

Para la práctica, el resultado indica que la manera de discutir la literatura importa tanto como la elección de los textos: incorporar tertulias con obras que representan el mundo andino al área de Comunicación es una vía de bajo costo para formar lectores capaces de interpretar, situar ideológicamente y evaluar con argumentos un relato. Para la investigación, las limitaciones señaladas —asignación no probabilística, mediación medida en el mismo momento que el resultado y empatía por autoinforme— definen la agenda inmediata: aleatorizar aulas, ampliar las mediciones y contrastar si el patrón se reproduce con otros corpus y en otros contextos.

14.7 REFERENCIAS BIBLIOGRÁFICAS

Al-Afifi, M., Nygren, T., Rasmusson, M., & Tväråna, M. (2026). Intercultural competence in the classroom: a systematic review of classroom-based research on teaching and learning practices for developing intercultural competence in different educational contexts. *Intercultural Education, 37*(2), 217-250. https://doi.org/10.1080/14675986.2025.2555782

Allotey, E., García, R., Villardón, L., & Soler, M. (2023). Transforming the educational experiences of marginalized students in Ghana through dialogic literary gatherings. *Humanities and Social Sciences*

Communications, 10(1). https://doi.org/10.1057/s41599-023-01801-z

Álvarez, G., García, R., Khalfaoui, A., Santiago, M., & Flecha, R. (2023). Preventing bullying of students with special educational needs through dialogic gatherings: a case study in elementary education. *Humanities and Social Sciences Communications, 10*(1). https://doi.org/10.1057/s41599-023-02470-8

Arias, E., Bos, M., Giambruno, C., & Zoido, P. (2023). *PISA in Latin America and the Caribbean 2022: Basic facts about PISA*. Inter-American Development Bank. https://doi.org/10.18235/0005308

Baki, Y. (2025). The Impact of Critical Listening and Critical Reading on Critical Thinking. *Behavioral Sciences, 15*(1), 34. https://doi.org/10.3390/bs15010034

Barak, M., & Lefstein, A. (2022). Opening Texts for Discussion: Developing Dialogic Reading Stances. *Reading Research Quarterly, 57*(2), 449-468. https://doi.org/10.1002/rrq.413

Chávez, O. (2024). Interculturalidad y bilingüismo: entre el normativismo funcional y las posturas emergentes en el diseño curricular oficial de formación docente en educación intercultural bilingüe en el Perú. *Lengua y Sociedad, 23*(1), 103-130. https://doi.org/10.15381/lengsoc.v23i1.25928

Coleman, J. (2025). Reading Orientations: Reading Power Otherwise in Critical Literacy Research. *Reading Research Quarterly, 60*(2). https://doi.org/10.1002/rrq.614

Dodell, D., & Tamir, D. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology: General, 147*(11), 1713-1727. https://doi.org/10.1037/xge0000395

Falk, C., Vogel, T., Hammami, S., & Miočević, M. (2023). Multilevel mediation analysis in R: A comparison of bootstrap and Bayesian approaches. *Behavior Research Methods, 56*(2), 750-764. https://doi.org/10.3758/s13428-023-02079-4

Fernández, A., De Ibarra, A., Domínguez, A., & García, R. (2025). Progress in mentalizing ability among people with psychosis through dialogic literary gatherings. *BMC Psychology, 13*(1). https://doi.org/10.1186/s40359-025-02381-0

Fernández, A., Valls, R., Melgar, P., & Tellado, I. (2021). Enhancing Literacy and Communicative Skills of Students With Disabilities in Special Schools Through Dialogic Literary Gatherings. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.662639

Fernández, R., & García, S. (2025). The Impact of Dialogic Literary Gatherings on the Development of Receptive Skills in EFL Undergraduates. *Miscelánea: A Journal of English and American Studies, 72*, 45-68. https://doi.org/10.26754/ojs_misc/mj.202511384

Flora, D., Crone, G., & Bell, S. (2025). Effect Size Interpretation in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 32*(6), 1069-1076. https://doi.org/10.1080/10705511.2025.2459768

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Frei, A., Gasser, L., Egger, S., Dammert, Y., Preisig, D., Murphy, P., & Hascher, T. (2026). Promoting social perspective-taking performance in late childhood through literary discussions. *International Journal of Educational Research*, *139*, 103056. https://doi.org/10.1016/j.ijer.2026.103056

Fulton, K., Lin, T., & Newell, G. (2023). Dialogic literary argumentation and close reading: effects on high school students' literature-related argumentative writing and motivational beliefs. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1214773

Garcia, J., Corbetta, S., & Baronnet, B. (2023). Decolonizing education in Latin America: critical environmental and intercultural education as an indigenous pluriversal alternative. *British Journal of Sociology of Education*, *44*(8), 1394-1412. https://doi.org/10.1080/01425692.2023.2234088

Gebremariam, H., & Weldeyohannes, M. (2025). Cultivating Students' Reading Skills: Inspiring Reading Comprehension and Motivation Through Critical Reading Strategies. *SAGE Open*, *15*(3). https://doi.org/10.1177/21582440251367189

Girbés, S., Tellado, I., López, G., & De Botton, L. (2024). Promoting high-quality interactions among early childhood education minority students: a case study of dialogic literary gatherings. *Literacy*, *58*(2), 228-239. https://doi.org/10.1111/lit.12354

Hedges, L., & Hedberg, E. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87. https://doi.org/10.3102/0162373707299706

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hogan, E., & Sun, Y. (2025). The Association Between Classroom Dialogic Interaction and Student Reading Performance: A Mixed Methods Study of Teacher Stance, Discourse Moves, and Reading Achievement. *Reading Research Quarterly, 60*(2). https://doi.org/10.1002/rrq.70009

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, Q., Cheung, A., Xu, J., & Jiang, Y. (2025). Intercultural learning and adolescents' global competence: The mediating effect of self-efficacy. *International Journal of Intercultural Relations, 105*, 102143. https://doi.org/10.1016/j.ijintrel.2025.102143

Khaqan, S., Redondo, G., & Flecha, R. (2025). Improving Social Interactions Among Adult Carers of Children with Disabilities Through Dialogic Literary Gatherings: A Case Study from Pakistan. *Behavioral Sciences, 15*(4), 509. https://doi.org/10.3390/bs15040509

Kim, J., & Thoemmes, F. (2025). Causal Assumptions of the Two-Wave Longitudinal Mediation Model. *Multivariate Behavioral Research*, *60*(1), 17-18. https://doi.org/10.1080/00273171.2025.2443360

Kiziltaş, Y., & Kultas, E. (2026). Why Disadvantaged Primary School Students in Rural Areas Have Lower Reading Skills? A Critical Assessment. *Reading Psychology*, *47*(2), 177-211. https://doi.org/10.1080/02702711.2025.2551555

Lenhart, J., Dangel, J., & Richter, T. (2022). The relationship between lifetime book reading and empathy in adolescents: Examining transportability as a moderator. *Psychology of Aesthetics, Creativity, and the Arts*, *16*(4), 679-693. https://doi.org/10.1037/aca0000341

Liñán, H., Sano, J., Paucar, C., & Tafur, L. (2023). Intercultural Bilingual Education (EIB) in Peru. *Research in Subject-matter Teaching and Learning (RISTAL)*, *6*(1), 84-95. https://doi.org/10.2478/ristal-2023-0006

Luong, R., & Flake, J. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*, *28*(4), 905-924. https://doi.org/10.1037/met0000441

Ma, L. (2025). Effects of literature circles activity on reading comprehension of L2 English learners: a meta-analysis. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-04695-1

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling.

Psychological Methods, 1(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Mackinnon, S., Curtis, R., & O'Connor, R. (2022). Tutorial in Longitudinal Measurement Invariance and Cross-lagged Panel Models Using Lavaan. *Meta-Psychology, 6.* https://doi.org/10.15626/mp.2020.2595

Marino, M., Flores, A., Rivas, H., & Vásquez, P. (2023). Myth and Identity: A Compilation of Oral Traditions in a University Context from Peru. *Journal of Intercultural Studies, 44*(6), 833-848. https://doi.org/10.1080/07256868.2023.2229258

Mekuria, A., Bushisho, E., & Wubshet, H. (2024). The effects of reading strategy training on students' reading strategy use and critical reading ability in EFL reading classes. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186x.2024.2310444

Moreano, G., Zegarra, C., & Rivera, Y. (2026). New intercultural bilingual education teacher preparation curriculum: Opportunities and challenges to reduce educational inequities in Peru. *Education Policy Analysis Archives, 34.* https://doi.org/10.14507/epaa.34.9008

Moshagen, M., & Bader, M. (2023). semPower: General power analysis for structural equation models. *Behavior Research Methods, 56*(4), 2901-2922. https://doi.org/10.3758/s13428-023-02254-7

Moulavinafchi, A., Oroskhan, M., & Alrammahi, M. (2026). Exploring the Predictive and Mediating Roles of Reading Motivation, Critical Reading Intention, and Reading Self-Efficacy on Reading Comprehension: A SEM Approach. *Reading Psychology, 47*(2), 352-387. https://doi.org/10.1080/02702711.2025.2589758

Murphy, P., Wilkinson, I., Soter, A., Hennessey, M., & Alexander, J. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740-764. https://doi.org/10.1037/a0015576

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Peng, P., Wang, W., Filderman, M., Zhang, W., & Lin, L. (2024). The Active Ingredient in Reading Comprehension Strategy Intervention for Struggling Readers: A Bayesian Network Meta-analysis. *Review of Educational Research*, *94*(2), 228-267. https://doi.org/10.3102/00346543231171345

Rezende, J., & Shigaeff, N. (2023). The effects of reading and watching fiction on the development of social cognition: a systematic review. *Dementia & Neuropsychologia, 17*. https://doi.org/10.1590/1980-5764-dn-2023-0066

Rice, M., Lambright, K., & Wijekumar, K. (2024). Professional Development in Reading Comprehension: A Meta-analysis of the Effects on Teachers and Students. *Reading Research Quarterly*, *59*(3), 424-447. https://doi.org/10.1002/rrq.546

Roca, E., Zubiri, H., León, S., & Aubert, A. (2024). Perceived Improvement of Literacy Skills of Students with and Without Special Educational Needs Through Dialogic Literary Gatherings. *Disabilities, 4*(4), 1030-1043. https://doi.org/10.3390/disabilities4040064

Ruiz, L., Soler, M., Racionero, S., & Padrós, M. (2023). Dialogic literary gatherings: A systematic review of evidence to overcome social and educational inequalities. *Educational Research Review, 39*, 100534. https://doi.org/10.1016/j.edurev.2023.100534

Salceda, M., Vidu, A., Aubert, A., & Padros, M. (2022). Dialogic literary gatherings in out-of-home care to overcome educational inequalities by improving school academic performance. *Children and Youth Services Review, 133*, 106368. https://doi.org/10.1016/j.childyouth.2022.106368

Santi, E., Cebula, K., & McGeown, S. (2025). Reading and Empathy: Qualitative insights into adolescents' experiences with fiction books. *Literacy, 59*(3), 304-313. https://doi.org/10.1111/lit.70006

Santiago, M., García, R., Zubiri, H., & López, G. (2023). Inclusion of L2 (Basque) learners in Dialogic Literary Gatherings in a linguistically diverse context. *Language Teaching Research, 27*(6), 1532-1551. https://doi.org/10.1177/1362168821994142

Schoemann, A., Boulton, A., & Short, S. (2017). Determining Power and Sample Size for Simple and Complex Mediation Models. *Social Psychological and Personality Science, 8*(4), 379-386. https://doi.org/10.1177/1948550617715068

Sullón, K. (2024). El paisaje lingüístico en las escuelas de Educación Intercultural Bilingüe del Perú. *Revista Iberoamericana de Educación, 96*(1), 83-95. https://doi.org/10.35362/rie9616436

Tang, M., Lau, K., & Du, Y. (2026). Effects and moderators of dialogic reading on children's reading literacy: A three-level meta-analysis

on studies from 2000 to 2025. *International Journal of Educational Research, 137,* 102963. https://doi.org/10.1016/j.ijer.2026.102963

Tibbe, T., & Montoya, A. (2022). Correcting the Bias Correction for the Bootstrap Confidence Interval in Mediation Analysis. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.810258

Tong, Y., & Ding, Y. (2025). Productive classroom dialogue and its association with student achievement in knowledge-building environments. *Language and Education, 39*(1), 232-251. https://doi.org/10.1080/09500782.2024.2323207

Valente, M., Georgeson, A., & Gonzalez, O. (2021). Clarifying the Implicit Assumptions of Two-Wave Mediation Models via the Latent Change Score Specification: An Evaluation of Model Fit Indices. *Frontiers in Psychology, 12.* https://doi.org/10.3389/fpsyg.2021.709198

Walsh, C. (2023). On Justice, Pedagogy, and Decolonial(Izing) Praxis. *Educational Theory, 73*(4), 511-529. https://doi.org/10.1111/edth.12592

Wang, Y., & Rhemtulla, M. (2021). Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1). https://doi.org/10.1177/2515245920918253

Wimmer, L., Currie, G., Friend, S., & Ferguson, H. (2021). The effects of reading narrative fiction on social and moral cognition. *Scientific Study of Literature, 11*(2), 223-265. https://doi.org/10.1075/ssol.21010.wim

Xie, W., & Lin, X. (2025). Are dialogic teaching-and-learning programs effective in pre-primary and primary schools? A meta-analysis. *Teaching and Teacher Education*, *153*, 104823. https://doi.org/10.1016/j.tate.2024.104823

Zsögön, M. (2025). Between assimilation and recognition of Indigenous Peoples: the role of intercultural bilingual education in Latin America. *International Journal of Social Pedagogy*, *14*(1). https://doi.org/10.14324/111.444.ijsp.2025.v14.x.002

CAPÍTULO XV

ESTRATEGIAS LECTORAS INTERCULTURALES BILINGÜES Y COMPRENSIÓN INFERENCIAL EN AULAS PRIMARIAS QUECHUA–CASTELLANO: UN ESTUDIO CUASIEXPERIMENTAL POR CONGLOMERADOS EN LOS ANDES PERUANOS

Ibeth Yusely Garcia Puma

Universidad Nacional de San Antonio Abad del Cusco

Resumen

La inferencia distingue a quienes construyen un modelo coherente del texto de quienes recuperan proposiciones aisladas, y es una dificultad persistente para el alumnado indígena que aprende a leer en una lengua escolar dominante. Se evaluó un programa intercultural bilingüe de lectura de 14 semanas que combinó la enseñanza explícita de inferencias con el uso sistemático del quechua y del conocimiento comunitario en lecciones en castellano. Se aplicó un diseño cuasiexperimental por conglomerados con pretest, postest y comparación no equivalente en 62 aulas multigrado de quinto y sexto grado del sur andino peruano (N = 990 estudiantes). En un análisis de covarianza multinivel, el alumnado del programa superó al de comparación en 3,15 puntos sobre 30 (IC 95 % [2,32; 3,98]; $p < 0,001$; g ajustada = 0,58). Los efectos fueron significativos en cada tipo de inferencia tras la corrección de Holm, mientras que el efecto en la comprensión literal fue pequeño y no significativo (0,27 puntos; $p = 0,124$). El nivel satisfactorio lo alcanzó el 67,5 % del alumnado del programa y el 41,1 % del de comparación (OR ajustada = 5,25), y la estimación fue estable en siete especificaciones alternativas. Los hallazgos sugieren que anclar la enseñanza de inferencias en la lengua y el conocimiento comunitario del alumnado mejora los procesos inferenciales que persigue.

Palabras clave: comprensión inferencial; educación intercultural bilingüe; translenguaje; quechua; estrategias de lectura; cuasiexperimento por conglomerados; análisis multinivel

Abstract

Inference making distinguishes readers who build a coherent situation model from those who retrieve isolated propositions, and it remains a persistent difficulty for Indigenous pupils who learn to read in a dominant school language. This study evaluated a 14-week intercultural bilingual reading programme that combined explicit inference instruction with the systematic use of Quechua and community knowledge in Spanish-medium lessons. A cluster quasi-experimental

pretest–posttest design with a nonequivalent comparison group was implemented in 62 multigrade fifth- and sixth-grade classrooms in the southern Peruvian Andes (31 per arm; N = 990 pupils). A 40-item test measured inferential and literal comprehension. In a multilevel analysis of covariance, programme pupils outperformed comparison pupils by 3.15 points on a 30-point scale (95% CI [2.32, 3.98]; p < .001; adjusted g = 0.58). Effects were significant for each inference type after Holm correction, whereas the effect on literal comprehension was small and not significant (0.27 points; p = .124). The satisfactory level was reached by 67.5% of programme pupils and 41.1% of comparison pupils (adjusted OR = 5.25), and the estimate was stable across seven alternative specifications. The findings suggest that anchoring inference instruction in pupils' home language and community knowledge improves the inferential processes it targets, and they support treating linguistic and cultural repertoires as cognitive resources for comprehension in Indigenous and minoritised bilingual education systems.

Keywords: *inferential comprehension; intercultural bilingual education; translanguaging; Quechua; reading strategies; cluster quasi-experiment; multilevel analysis*

15.1 INTRODUCCIÓN

La comprensión lectora exige que el lector vaya más allá de lo que el texto enuncia: que enlace oraciones mediante puentes causales y referenciales, que infiera el significado de las palabras desconocidas a partir del contexto y que elabore y evalúe el texto a la luz de su conocimiento previo (Duke et al., 2021; Rice et al., 2024). Estos procesos inferenciales permiten construir un modelo de la situación en lugar de una lista de proposiciones, y constituyen una dificultad persistente en los grados superiores de primaria y en la educación media para quienes presentan problemas de comprensión (Vaughn y Barnes, 2023). La evidencia metaanalítica indica que la inferencia es enseñable: en 25 estudios de los grados K–12, su instrucción mejoró la comprensión general (d = 0,58) y la inferencial (d = 0,68), con ganancias adicionales en la comprensión literal de los lectores menos hábiles (Elleman, 2017). Se han evaluado además programas de

enseñanza de inferencias en primaria (McDonald et al., 2021; Van Zeijts et al., 2024), un ensayo aleatorizado con tutores en secundaria produjo efectos significativos sobre los tipos de inferencia enseñados (g = 0,40; Barnes et al., 2024), y un metaanálisis en red ha examinado los componentes activos de las estrategias para lectores con dificultades (Peng et al., 2024).

Esta literatura presenta dos puntos ciegos que limitan su valor para la educación de la infancia indígena y minorizada. El primero es geográfico y lingüístico: el metaanálisis en red citado se apoya en muestras mayoritariamente de países anglófonos (Peng et al., 2024), los ensayos recientes sobre programas de comprensión se han dirigido a alumnado en riesgo o a escuelas desfavorecidas, no a poblaciones indígenas bilingües (Nergård et al., 2026; O'Hare et al., 2023), y sin embargo muchos niños con menor comprensión aprenden a leer en una lengua escolar dominante que no es la de su hogar (Wawire et al., 2024). El segundo es conceptual: la enseñanza de inferencias suele diseñarse como una rutina cognitiva desligada del conocimiento que los niños aportan al texto, aunque la generación de inferencias se nutre directamente del conocimiento previo (Smith et al., 2021). Para un niño quechuahablante que lee un texto en castellano, el conocimiento necesario para tender un puente causal o elaborar los motivos de un personaje suele estar codificado en quechua y en prácticas comunitarias que los textos escolares rara vez invocan.

La educación intercultural bilingüe (EIB) es la respuesta de varios Estados latinoamericanos a este desajuste. En el Perú es el modelo oficial para las escuelas de comunidades con lenguas indígenas, y el currículo nacional incorpora las lenguas originarias y los saberes locales en la educación básica regular (Liñán et al., 2023). Su implementación, sin embargo, es desigual: la formación docente para la EIB se ha reformado solo recientemente, con oportunidades y desafíos para reducir las inequidades educativas (Chávez et al., 2026; Moreano et al., 2026); programas comparables en Ecuador muestran prácticas de aula divergentes en el uso de la lengua indígena (Limerick, 2023), y los estudios sobre alfabetización temprana en escuelas EIB peruanas documentan perfiles diversos entre la infancia indígena (Guevara et al., 2026). Sobre todo, hay escasa evidencia experimental o

cuasiexperimental acerca de si las pedagogías EIB producen ganancias medibles en las habilidades de comprensión de orden superior que el currículo nacional exige en los grados superiores de primaria.

El translenguaje ofrece un mecanismo que vincula ambas literaturas. La pedagogía translingüe trata el repertorio lingüístico completo del bilingüe como recurso para construir significado, y el trabajo experimental ha examinado su efecto sobre la comprensión lectora en segunda lengua cuando el alumnado puede emplear su primera lengua para discutir e interpretar los textos (Qureshi y Aljanadbah, 2022). Falta, sin embargo, someter a prueba si un programa centrado en la inferencia que movilice deliberadamente la lengua materna y el conocimiento comunitario mejora los procesos inferenciales concretos que la instrucción estratégica general persigue.

Este estudio aborda esa laguna. Evalúa un programa intercultural bilingüe de lectura inferencial en aulas primarias quechua–castellano del sur andino peruano mediante un diseño cuasiexperimental por conglomerados con medidas pretest y postest, modelos multinivel que respetan el anidamiento del alumnado en las aulas y una medida que separa tres tipos de inferencia de la comprensión literal. Aunque el escenario es andino, la pregunta es general: si los beneficios cognitivos de la enseñanza de inferencias dependen de los recursos lingüísticos y culturales que se permite al lector aportar al texto. La respuesta interesa al diseño de la instrucción lectora en cualquier sistema donde la infancia aprende a leer en una lengua distinta de la propia, y aporta una estimación basada en el diseño a una literatura que, para los contextos bilingües indígenas, ha dependido de evidencia descriptiva y cualitativa.

15.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

15.2.1 Inferencia y modelo de la situación

Los modelos de comprensión distinguen la base del texto, que representa lo que este dice, del modelo de la situación, que integra el texto con el

conocimiento del lector (Duke et al., 2021). Las inferencias son las operaciones que construyen este último. Las inferencias léxicas recuperan el significado de palabras desconocidas a partir del contexto; las de puente causal conectan sucesos y estados que el texto presenta por separado; y las elaborativo-evaluativas extienden el texto con predicciones, motivos y juicios apoyados en el conocimiento previo (Rice et al., 2024; Van Zeijts et al., 2024). Los tres tipos se nutren del conocimiento en proporciones distintas: las léxicas y las de puente están más acotadas por el texto, mientras que las elaborativas dependen en mayor medida de lo que el lector ya sabe (Smith et al., 2021). De ahí que la tipología sea relevante para la evaluación, pues cada tipo puede responder de modo diferenciado a una enseñanza que active el conocimiento comunitario. La instrucción que explicita el proceso inferencial mediante modelado, formulación de preguntas y organizadores gráficos mejora la inferencia en distintas edades y niveles de habilidad (Elleman, 2017; McDonald et al., 2021).

15.2.2 Recursos bilingües y culturales como conocimiento inferencial

Para el niño bilingüe que aprende a leer en su segunda lengua, el conocimiento necesario para inferir se distribuye entre dos lenguas. Una inferencia que el niño haría con facilidad en su lengua materna puede fracasar en la lengua escolar por falta de acceso al conocimiento pertinente y no por carencia de habilidad inferencial. La pedagogía intercultural bilingüe añade un segundo recurso: textos y conversaciones que movilizan el saber comunitario aportan los esquemas que requieren las inferencias causales y elaborativas (Liñán et al., 2023). El translenguaje permite que ambos recursos entren en la lección sin abandonar el texto en castellano, pues el alumnado puede discutir, parafrasear y justificar sus inferencias en quechua mientras el texto y las respuestas escritas permanecen en castellano (Qureshi y Aljanadbah, 2022). Si esta lectura es correcta, la enseñanza de inferencias que ignore el repertorio lingüístico y cultural del alumnado subestimará su potencial, y la que lo movilice deberá mostrar ganancias específicas en los procesos inferenciales, no en la mera recuperación de información.

15.2.3 Hipótesis

El marco conduce a tres hipótesis. Dado que el programa explicita la inferencia y la conecta con conocimientos que el alumnado posee en quechua y en las prácticas comunitarias, se espera que el alumnado del programa supere al de comparación en comprensión inferencial al final de la intervención, tras controlar su nivel inicial (H1). Dado que el programa enseña de forma explícita los tres tipos de inferencia, se espera el efecto en cada uno de ellos: léxica, de puente causal y elaborativo-evaluativa (H2). Dado que el programa se dirige a la inferencia y no a la recuperación de información explícita, se espera que su efecto sobre la comprensión inferencial sea mayor que su efecto sobre la comprensión literal (H3).

15.3 MATERIALES Y MÉTODOS

15.3.1 Diseño y participantes

Se empleó un diseño cuasiexperimental por conglomerados, con pretest, postest y grupo de comparación no equivalente, en el que las aulas fueron las unidades de asignación. El estudio se realizó en 2025 en escuelas públicas primarias interculturales bilingües de fortalecimiento de Apurímac y Cusco, con comunidades rurales de lengua del hogar quechua y quinto y sexto grado en secciones multigrado. Las aulas del programa correspondían a las escuelas incorporadas en su primer ciclo de implementación; las de comparación, a escuelas de las mismas provincias, tipo y organización que mantuvieron el currículo regular de comunicación. La asignación no fue aleatoria, y la muestra, reclutada mediante las unidades locales de gestión educativa, fue de conveniencia, sin pretensión de generalización estadística.

El tamaño muestral se calculó a priori con $\alpha = 0{,}05$ bilateral, potencia de 0,80 y un efecto mínimo de interés de $d = 0{,}50$ en comprensión inferencial, inferior a las estimaciones metaanalíticas. Para el análisis de covarianza se aplicó el factor $(1 - \rho^2)$ con $\rho = 0{,}50$ (Borm et al., 2007), y los requerimientos se multiplicaron por el efecto de diseño $1 + (m - 1)ICC$,

con m = 15 e ICC = 0,20 tomado del rango compilado para el rendimiento lector (Hedges y Hedberg, 2007), es decir, 3,80. La prueba más exigente fue el contraste de especificidad de H3 (d = 0,50 frente a 0,20; correlación de 0,55 entre resultados), resuelto mediante simulación de Monte Carlo (1.000 réplicas por diseño): la potencia fue de 0,779 con 56 aulas y de 0,833 con 62, de modo que se requirieron 62 aulas (31 por brazo) y 930 estudiantes. La simulación del conjunto de pruebas arrojó potencias de 0,805 (H3), 0,991 (H1), 0,973 (interacción grupo × tiempo) y entre 0,984 y 0,990 (H2, con α ajustado por Holm). Se inscribieron 1.029 estudiantes, y la muestra analítica de 990 con ambas mediciones (498 y 492) superó el requerimiento.

15.3.2 Programa de lectura inferencial intercultural bilingüe

El programa duró 14 semanas (abril–julio de 2025), con tres sesiones semanales de 60 minutos que sustituyeron parte de las lecciones de comunicación. Cada sesión comprendía cuatro fases: activación del conocimiento comunitario en un intercambio oral en quechua; lectura en voz alta de un texto en castellano con modelado de una inferencia mediante pensamiento en voz alta bilingüe; trabajo en pares o grupos sobre textos nuevos, con discusión en quechua o castellano y registro en castellano en organizadores gráficos bilingües que distinguían inferencias léxicas, causales y de predicción-juicio; y comparación grupal de inferencias y justificaciones. La mitad de los 42 textos eran narrativas y textos expositivos andinos; la otra mitad, materiales nacionales. Los docentes recibieron 24 horas de formación y visitas mensuales de acompañamiento, y la fidelidad, registrada en bitácoras y verificada en las visitas, promedió el 87,8 % (rango 68,0 %–99,0 %). La comparación siguió el programa regular con igual tiempo semanal.

15.3.3 Medidas y procedimiento

La comprensión se midió con una prueba en castellano de 40 ítems dicotómicos sobre seis pasajes (tres narrativos y tres expositivos, dos con contenido andino), en cuatro subescalas de diez ítems: inferencia léxica,

de puente causal, elaborativo-evaluativa y comprensión literal (Tabla 15-1); la comprensión inferencial es la suma de las tres primeras (0–30). Se piloteó con 64 estudiantes ajenos al estudio y la administraron evaluadores capacitados, distintos de los docentes, en una sesión de 80 minutos por ola, con instrucciones en quechua y castellano, el mismo formulario y distinto orden de pasajes. La fiabilidad se estimó con KR-20 y ω de McDonald. Sexo, grado y lengua dominante procedieron de registros escolares y de una lista de cotejo docente. El pretest se aplicó en abril y el postest en agosto de 2025, con datos codificados.

Tabla 15-1

Variables, instrumentos y puntuación

Variable	Rol	Instrumento / fuente	Ítems	Puntuación y rango
Programa	Factor (nivel aula)	Asignación de aulas	—	1 = aula del programa, 0 = aula de comparación
Inferencia léxica	Resultado (H2)	Prueba de comprensión inferencial y literal, subescala LI	10	Ítems dicotómicos; suma 0–10
Inferencia de puente causal	Resultado (H2)	Misma prueba, subescala CI		
Inferencia elaborativo-evaluativa	Resultado (H2)	Misma prueba, subescala EI		
Comprensión inferencial	Resultado primario (H1, H3)	Suma de LI, CI y EI	30	0–30
Comprensión literal	Resultado de especificidad (H3)	Misma prueba, subescala LIT	10	Ítems dicotómicos; suma 0–10
Puntajes del pretest	Covariables	Misma prueba en la línea base	40	Centrados en la media muestral
Grado, sexo, lengua dominante		Registros escolares; lista de cotejo docente	—	Sexto grado = 1; niña = 1; quechua-dominante = 1
Fidelidad de implementación	Dosis (brazo del programa)	Bitácoras verificadas en visitas de acompañamiento	—	Porcentaje de sesiones con todas las fases

Nota: LI = inferencia léxica; CI = inferencia de puente causal; EI = inferencia elaborativo-evaluativa; LIT = comprensión literal.

15.3.4 Análisis de datos

La equivalencia basal se evaluó con diferencias estandarizadas (g de Hedges; índice de Cox) y pruebas que consideran el agrupamiento; siguiendo al What Works Clearinghouse (2022), diferencias hasta 0,05 se consideraron equivalentes y entre 0,05 y 0,25 requirieron ajuste, y se calculó el desgaste global y diferencial. El análisis primario (H1) fue un análisis de covarianza multinivel del puntaje inferencial postest, con intercepto aleatorio por aula y máxima verosimilitud restringida, y con programa, pretest centrado, grado, sexo y lengua dominante como efectos fijos; se examinaron los supuestos y la interacción programa × pretest (con región de Johnson–Neyman), y el cambio se analizó con modelos multinivel de la ganancia. Para H2 se estimó el mismo modelo por subescala con corrección de Holm y un MANCOVA a nivel de aula (J = 62); para H3, los puntajes inferencial y literal se estandarizaron con la media y la DE del pretest y se modelaron con ecuaciones de estimación generalizadas (GEE) agrupadas por aula. Los tamaños del efecto fueron d_ppc2 (Morris, 2008) y g ajustada, con IC por bootstrap de aulas (2.000 remuestreos). El nivel satisfactorio (≥ 21 de 30) se comparó con una GEE logística y la dosis-respuesta se examinó con la fidelidad. La robustez se evaluó con siete especificaciones alternativas, entre ellas el análisis por aula, la imputación múltiple (20 imputaciones) y una prueba de permutación (2.000). Se empleó Python (statsmodels y SciPy), con $\alpha = 0{,}05$ bilateral.

15.4 RESULTADOS

15.4.1 Muestra, desgaste y equivalencia basal

Las 62 aulas inscribieron 1.029 estudiantes (518 en el programa y 511 en la comparación), con entre 12 y 20 estudiantes por aula (media 16,6). Treinta y nueve estudiantes (3,8 %) no rindieron el postest: 20 en el programa (3,9 %) y 19 en la comparación (3,7 %), un desgaste diferencial de 0,1 puntos porcentuales, dentro del límite de bajo desgaste del What Works Clearinghouse (2022). La muestra analítica comprendió 990 estudiantes

(498 y 492). Como muestra la Tabla 15-2, los brazos fueron equivalentes en edad y sexo (|ES| ≤ 0,05); las diferencias en grado, lengua dominante y los cinco puntajes del pretest se situaron entre 0,04 y 0,10 en valor absoluto, rango que exige ajuste estadístico, y ninguna fue significativa al considerar el agrupamiento (todas p > 0,22). Todos los análisis siguientes se ajustan por el pretest, el grado, el sexo y la lengua dominante.

Tabla 15-2

Características basales y equivalencia de la muestra analítica (N = 990; 62 aulas)

Característica		Programa (n = 498)	Comparación (n = 492)	Diferencia estandarizada	p (agrupada)
Edad (años), M (DE)		11,13 (0,71)	11,15 (0,69)	−0,032	0,626
Niñas, %		48,8	47,8	0,025	0,719
Sexto grado, %		49,0	51,4	−0,059	0,444
Quechua-dominante, %		62,9	59,3	0,089	0,229
Inferencia pretest (0–10), M (DE)	léxica	4,46 (2,84)	4,29 (2,85)	0,062	0,424
	de puente causal	6,12 (2,57)	6,02 (2,62)	0,040	0,562
	elaborativo-evaluativa,	5,85 (2,35)	5,67 (2,35)	0,076	0,326
Comprensión pretest (0–30), M (DE)	inferencial,	16,43 (6,87)	15,97 (6,96)	0,066	0,397
	literal	5,94 (2,46)	5,69 (2,38)	0,102	0,267

Nota: Diferencia estandarizada: g de Hedges para variables continuas e índice de Cox para proporciones (programa menos comparación). Valores p de modelos lineales mixtos con intercepto aleatorio por aula (variables continuas) y de ecuaciones de estimación generalizadas con correlación intercambiable por aula (proporciones).

15.4.2 Propiedades psicométricas de la medida

La fiabilidad de las subescalas osciló entre 0,704 y 0,795 en el pretest y entre 0,712 y 0,806 en el postest (KR-20), con coeficientes ω casi idénticos, y el total inferencial alcanzó 0,897 y 0,858 (Tabla 15-3). Las dificultades de los ítems cubrieron el rango de 0,17 a 0,90 y todas las correlaciones ítem–total corregidas fueron de al menos 0,26. La menor fiabilidad postest del

total inferencial coincidió con una varianza reducida tras la instrucción, al aproximarse al techo los ítems más fáciles en el brazo del programa.

Tabla 15-3

Fiabilidad y estadísticos de los ítems de la prueba de comprensión (N = 990)

Ola	Escala	Ítems	KR-20	ω	Dificultad del ítem (mín.–máx.)	r ítem–total corregida (mín.–máx.)
Pretest	Inferencia léxica	10	0,795	0,797	0,29–0,66	0,40–0,57
	Inferencia de puente causal		0,761	0,763	0,26–0,77	0,33–0,54
	Inferencia elaborativo-evaluativa		0,715	0,717	0,17–0,83	0,33–0,46
	Comprensión inferencial (total)	30	0,897	0,897	0,17–0,83	0,35–0,57
	Comprensión literal	10	0,704	0,704	0,29–0,82	0,27–0,46
Postest	Inferencia léxica		0,806	0,807	0,46–0,80	0,41–0,55
	Inferencia de puente causal		0,771	0,781	0,45–0,88	0,36–0,52
	Inferencia elaborativo-evaluativa		0,712	0,721	0,30–0,90	0,30–0,42
	Comprensión inferencial (total)	30	0,858	0,860	0,30–0,90	0,26–0,50
	Comprensión literal	10	0,731	0,737	0,41–0,89	0,35–0,44

Nota: KR-20 = Kuder–Richardson 20; ω = omega de McDonald a partir de un modelo unifactorial de cada escala. La dificultad del ítem es la proporción de respuestas correctas.

15.4.3 Resultados descriptivos

Ambos brazos mejoraron entre mediciones, pero el del programa lo hizo más en todas las escalas inferenciales (Tabla 15-4). En el total inferencial, el alumnado del programa ganó 5,75 puntos (de 16,43 a 22,18) y el de comparación 2,79 (de 15,97 a 18,76). La mayor diferencia bruta de ganancias correspondió a la inferencia de puente causal, mientras que la comprensión literal cambió de modo similar en ambos brazos. Las correlaciones pretest–postest fluctuaron entre 0,484 y 0,696.

Tabla 15-4

Puntajes de pretest, postest y cambio por brazo (N = 990)

Escala (rango)	Programa M (DE)			Comparación M (DE)			r pre–post
	pretest	postest	cambio	pretest	postest	cambio	
Inferencia léxica (0–10)	4,46 (2,84)	6,54 (2,77)	2,08 (2,63)	4,29 (2,85)	5,43 (2,87)	1,15 (2,84)	0,529
Inferencia de puente causal (0–10)	6,12 (2,57)	8,12 (1,98)	2,00 (2,37)	6,02 (2,62)	6,77 (2,50)	0,76 (2,49)	0,490
Inferencia elaborativo-evaluativa (0–10)	5,85 (2,35)	7,53 (1,93)	1,68 (2,10)	5,67 (2,35)	6,55 (2,29)	0,88 (2,42)	0,484
Comprensión inferencial (0–30)	16,43 (6,87)	22,18 (5,12)	5,75 (4,58)	15,97 (6,96)	18,76 (5,86)	2,79 (5,07)	0,696
Comprensión literal (0–10)	5,94 (2,46)	7,03 (2,36)	1,09 (2,48)	5,69 (2,38)	6,65 (2,40)	0,96 (2,41)	0,484

Nota: Programa n = 498; comparación n = 492. Cambio = postest menos pretest.

15.4.4 Supuestos de los modelos y agrupamiento

Los residuos de nivel 1 de los análisis de covarianza fueron aproximadamente simétricos (asimetría entre –0,63 y –0,08; curtosis en exceso entre –0,33 y 0,32). La prueba de Brown–Forsythe indicó menor varianza residual en el brazo del programa para la inferencia de puente causal, la elaborativo-evaluativa y el total inferencial (todas $p < 0,001$) y para la inferencia léxica ($p = 0,023$); las estimaciones robustas a nivel de aula, por bootstrap y por permutación que se informan más adelante no dependen de la igualdad de varianzas (Tabla 14-5). La interacción programa × pretest no fue significativa para el total inferencial ($b = –0,020$; $p = 0,554$) ni para tres de las cuatro subescalas; sí lo fue para la inferencia de puente causal ($b = –0,120$; $p = 0,011$), donde el efecto fue de 1,61 puntos (IC 95 % [1,21; 2,01]) a una DE por debajo de la media del pretest y de 0,98 puntos (IC 95 % [0,59; 1,38]) a una DE por encima, y el análisis de Johnson–Neyman mostró un efecto significativo en todo el rango 0–10 de la subescala. Antes del ajuste por covariables, entre el 6,4 % y el 20,9 % de la varianza del postest se situó entre aulas, con efectos de diseño observados entre 1,96

y 4,13; tras el ajuste, la proporción atribuible al aula descendió a entre el 5,0 % y el 13,6 %.

Tabla 15-5

Verificación de supuestos y agrupamiento de los resultados del postest (N = 990; 62 aulas)

Resultado	Asimetría residual	Curtosis residual	F de Brown–Forsythe (p)	Programa × pretest b (p)	ICC incondicional	ICC condicional	Efecto de diseño
Inferencia léxica	−0,217	−0,332	5,190 (0,023)	0,033 (0,530)	0,090	0,072	2,34
Inferencia de puente causal	−0,632	0,308	24,011 (< 0,001)	−0,120 (0,011)	0,131	0,050	2,96
Inferencia elaborativo-evaluativa	−0,505	0,197	19,047 (< 0,001)	0,001 (0,979)	0,209	0,051	4,13
Comprensión inferencial (total)	−0,082	0,318	13,223 (< 0,001)	−0,020 (0,554)	0,176	0,136	3,64
Comprensión literal	−0,469	−0,289	0,250 (0,617)	−0,016 (0,774)	0,064	0,050	1,96

Nota: Residuos del análisis de covarianza multinivel de cada resultado (intercepto aleatorio por aula; pretest, grado, sexo y lengua dominante como covariables). La curtosis es la curtosis en exceso. Efecto de diseño = 1 + (m − 1) × ICC incondicional, con tamaño medio analítico de aula m = 15,97.

15.4.5 Efecto del programa sobre la comprensión inferencial (H1)

El análisis de covarianza multinivel mostró una ventaja del programa (Tabla 15-6). Con ajuste por pretest, grado, sexo y lengua dominante, el alumnado del programa puntuó 3,15 puntos más que el de comparación en la escala inferencial de 30 puntos (b = 3,154; EE = 0,424; IC 95 % [2,324; 3,984]; z = 7,45; p < 0,001), con medias ajustadas de 22,08 y 18,93. El efecto estandarizado ajustado fue g = 0,58 (IC 95 % por bootstrap de aulas [0,43; 0,74]). El pretest fue el predictor más fuerte (b = 0,569 por punto) y el alumnado de sexto grado puntuó 0,71 puntos más que el de quinto, mientras que el sexo y la lengua dominante no se asociaron con el postest.

Los efectos fijos explicaron el 56,2 % de la varianza del postest y el modelo completo el 62,1 %. Con la interacción programa × pretest, el efecto en la media del pretest (16,20 puntos) fue de 3,155 (EE = 0,423; p < 0,001), de 3,30 puntos (IC 95 % [2,34; 4,26]) a una DE por debajo y de 3,02 puntos (IC 95 % [2,07; 3,96]) a una DE por encima, y resultó significativo en todo el rango 0–30 del pretest (Figura 15-1).

Tabla 15-6

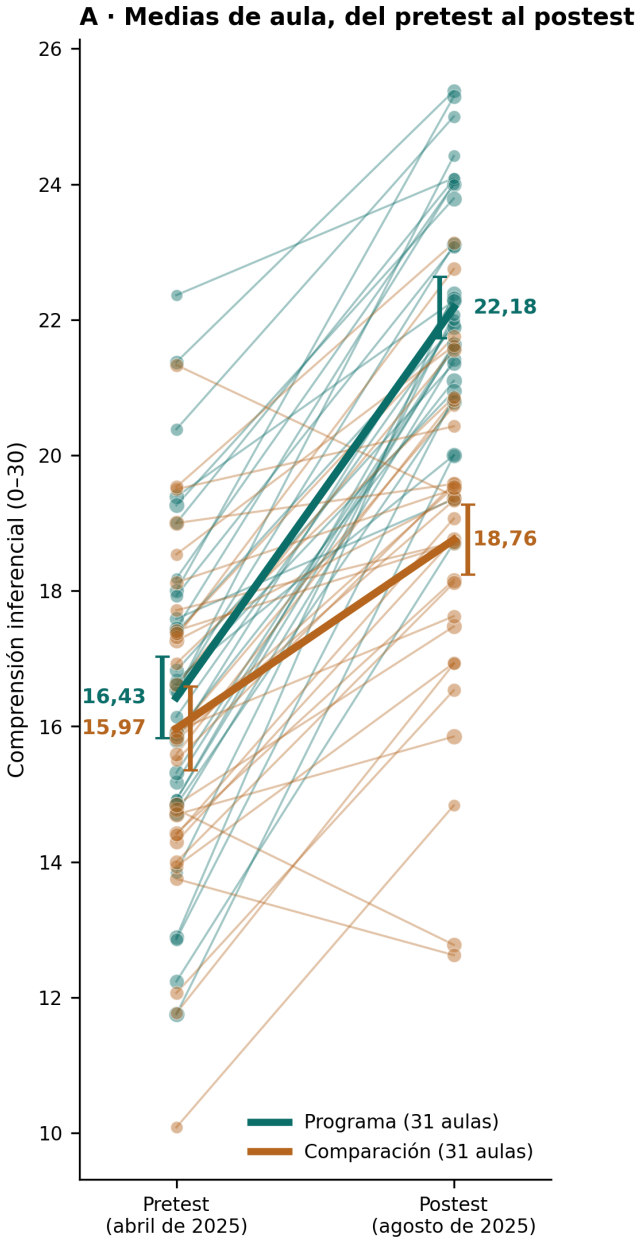
Análisis de covarianza multinivel de la comprensión inferencial del postest (N = 990; 62 aulas)

Término	b	EE	*IC 95 %*	z	p
Intercepto	18,493	0,371	[17,766; 19,219]	49,89	< 0,001
Programa (vs. comparación)	3,154	0,424	[2,324; 3,984]	7,45	< 0,001
Puntaje inferencial del pretest (centrado)	0,569	0,017	[0,535; 0,602]	33,13	< 0,001
Sexto grado (vs. quinto)	0,714	0,232	[0,260; 1,169]	3,08	0,002
Niña (vs. niño)	0,053	0,230	[–0,397; 0,504]	0,23	0,817
Quechua-dominante (vs. castellano-dominante)	0,080	0,237	[–0,385; 0,544]	0,34	0,736
Varianza del aula (τ_{00})	1,973	—	—	—	—
Varianza residual (σ^2)	12,586	—	—	—	—

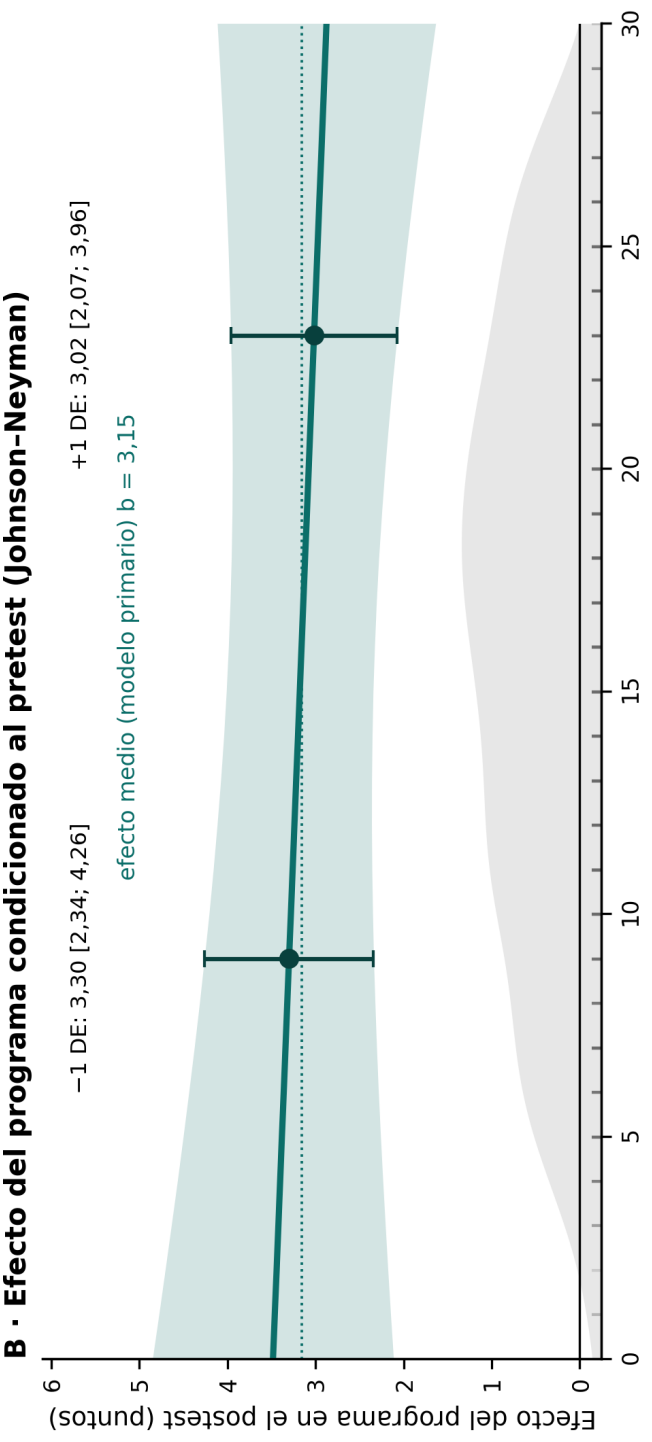
Nota: Máxima verosimilitud restringida. ICC condicional = 0,136; R^2 marginal = 0,562; R^2 condicional = 0,621. Medias ajustadas del postest: programa 22,08, comparación 18,93.

Figura 15-1

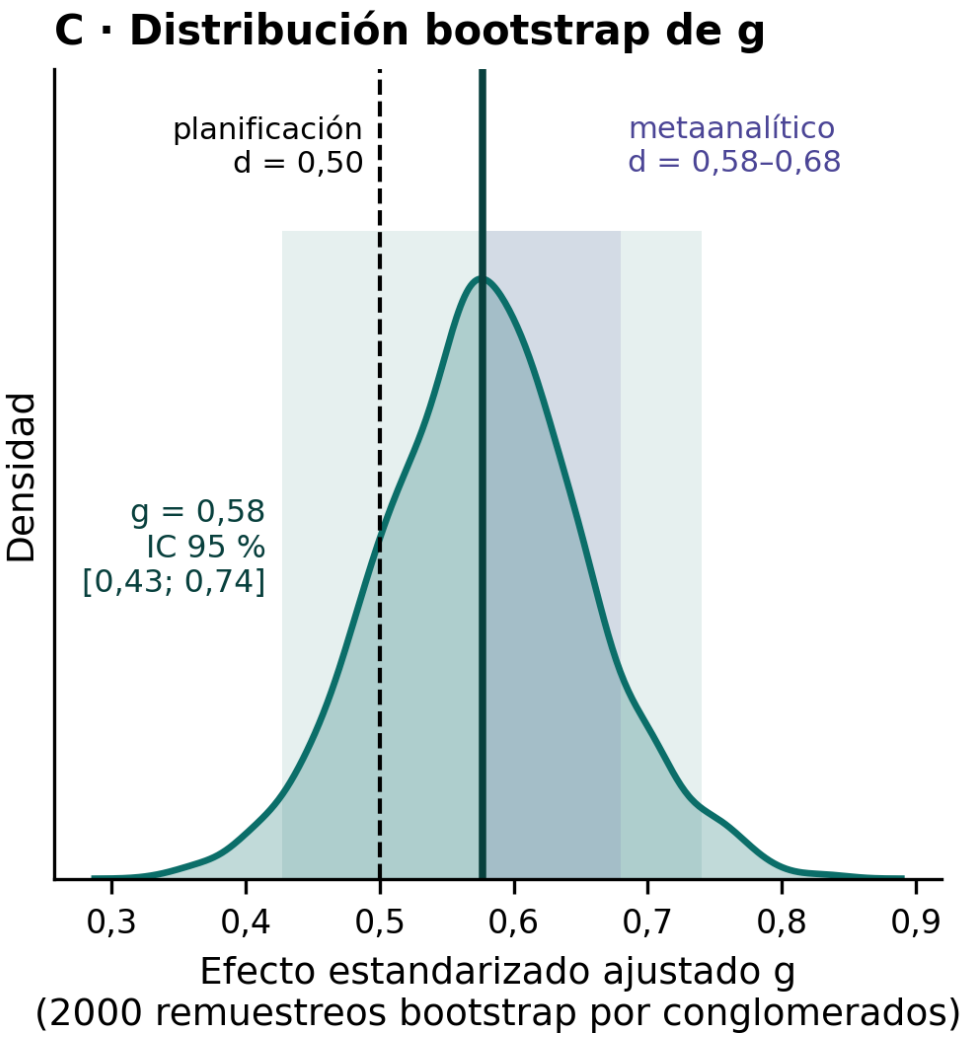
Efecto del programa en la comprensión inferencial (N = 990 alumnos en 62 aulas).



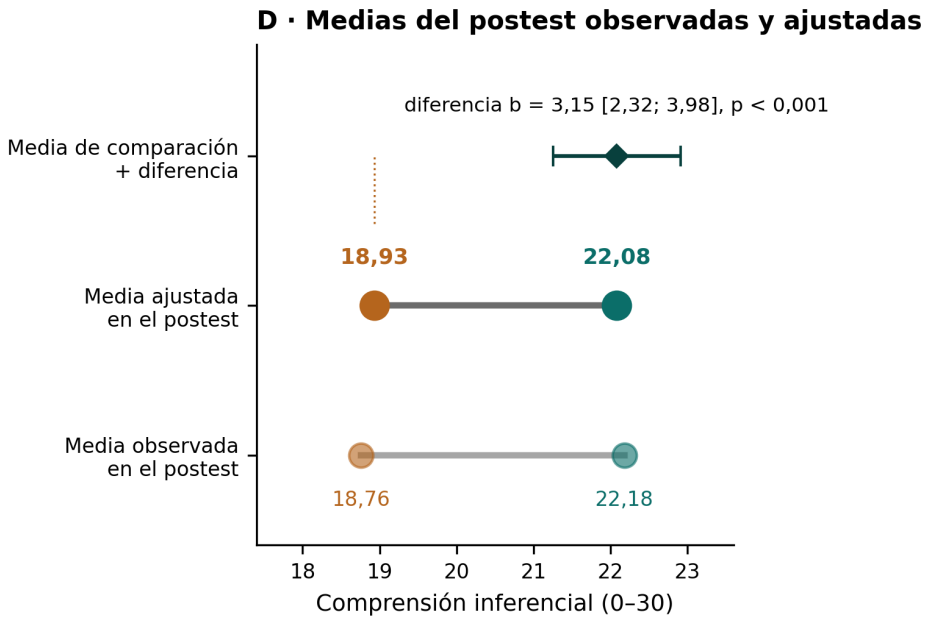
Nota. Panel A: medias de aula en comprensión inferencial (0–30) en el pretest y el postest; las líneas finas unen las dos mediciones de cada aula (área del marcador proporcional al tamaño del aula) y las gruesas unen las medias de cada brazo, con IC 95 % (programa, de 16,43 a 22,18; comparación, de 15,97 a 18,76).



Nota. Panel B: efecto del programa en la comprensión inferencial del postest en función de la puntuación del pretest, según el análisis de covarianza multinivel con la interacción programa × pretest (banda: IC 95 %; marcadores: efectos a una DE por debajo y por encima de la media del pretest, 3,30 [2,34; 4,26] y 3,02 [2,07; 3,96]; línea punteada: efecto del modelo primario, b = 3,15); la interacción no fue significativa (b = −0,020; p = 0,554) y el análisis de Johnson–Neyman mostró un efecto significativo en todo el rango 0–30; el área gris muestra la distribución de las puntuaciones del pretest.



Nota. Panel C: distribución del efecto estandarizado ajustado g en 2000 remuestreos bootstrap de aulas dentro de cada brazo (g = 0,58; IC 95 % [0,43; 0,74]), con el efecto de planificación d = 0,50 (línea discontinua) y el rango de estimaciones metaanalíticas de la enseñanza de inferencias, d = 0,58–0,68 (Elleman, 2017; banda sombreada).



Nota. Panel D: medias observadas del postest (comparación 18,76; programa 22,18), medias ajustadas del modelo primario (18,93 y 22,08) y diferencia ajustada b = 3,15; IC 95 % [2,32; 3,98]; p < 0,001.

15.4.6 Cambio entre el pretest y el postest

Los modelos multinivel de cambio reprodujeron el patrón de los análisis de covarianza (Tabla 15-7). La diferencia entre las ganancias medias de ambos brazos fue de 2,89 puntos en el total inferencial (IC 95 % [1,91; 3,88]; p < 0,001) y fue significativa en las tres subescalas inferenciales, mientras que en la comprensión literal fue cercana a cero (p = 0,520).

Tabla 15-7

Efectos grupo × tiempo: modelos multinivel de cambio pretest–postest (N = 990; 62 aulas)

Resultado		Cambio observado, programa	Cambio observado, comparación	Diferencia en el cambio b	EE	IC 95 %	p
Inferencia	léxica	2,08	1,15	0,904	0,254	[0,406; 1,401]	< 0,001
	de puente causal	2,00	0,76	1,212	0,199	[0,821; 1,603]	< 0,001
	elaborativo-evaluativa	1,68	0,88	0,794	0,167	[0,468; 1,121]	< 0,001
Comprensión	inferencial (total)	5,75	2,79	2,892	0,501	[1,909; 3,875]	< 0,001
	literal	1,09	0,96	0,131	0,203	[–0,268; 0,530]	0,520

Nota: Puntuación de ganancia (postest menos pretest) regresada sobre el indicador de programa, grado, sexo y lengua dominante, con intercepto aleatorio por aula (máxima verosimilitud restringida). La diferencia en el cambio es la interacción grupo × tiempo del diseño de dos olas.

15.4.7 Tipos de inferencia y especificidad (H2 y H3)

Los análisis de covarianza multinivel de las tres subescalas inferenciales mostraron efectos ajustados del programa de 1,01 puntos en la inferencia léxica, de 1,30 en la de puente causal y de 0,92 en la elaborativo-evaluativa, todos significativos tras la corrección de Holm (Tabla 15-8). El MANCOVA a nivel de aula indicó un efecto conjunto sobre las tres subescalas (V de Pillai = 0,535; F(3, 55) = 21,08; p < 0,001). El efecto sobre la comprensión literal fue de 0,27 puntos y no significativo (p = 0,124);

en la métrica estandarizada con el pretest, el efecto sobre la comprensión inferencial superó al efecto sobre la literal en 0,35 desviaciones estándar (IC 95 % [0,23; 0,47]; z = 5,60; p < 0,001).

Tabla 15-8

Efectos del programa por tipo de inferencia, prueba multivariada y contraste de especificidad

Análisis	Estimación	EE	IC 95 %	Estadístico de prueba	p	
					robusto	Holm
Inferencia léxica: diferencia ajustada (puntos)	1,008	0,220	[0,577; 1,439]	z = 4,58	< 0,001	< 0,001
Inferencia de puente causal: diferencia ajustada (puntos)	1,295	0,164	[0,973; 1,617]	z = 7,89	< 0,001	< 0,001
Inferencia elaborativo-evaluativa: diferencia ajustada (puntos)	0,916	0,157	[0,609; 1,223]	z = 5,85	< 0,001	< 0,001
Comprensión literal: diferencia ajustada (puntos)	0,272	0,177	[−0,075; 0,618]	z = 1,54	0,124	—
MANCOVA a nivel de aula de las tres subescalas inferenciales (J = 62)	V de Pillai = 0,535; Λ de Wilks = 0,465	—	—	F(3, 55) = 21,08	< 0,001	—
Especificidad: programa × tipo de resultado (inferencial vs. literal, unidades de DE del pretest)	0,351	0,063	[0,228; 0,474]	z = 5,60	< 0,001	—

Nota: Filas 1–4: análisis de covarianza multinivel con intercepto aleatorio por aula, con ajuste por el pretest correspondiente, grado, sexo y lengua dominante (N = 990). Corrección de Holm aplicada a las tres subescalas inferenciales. Fila 5: medias de aula con las medias de las tres subescalas del pretest como covariables. Fila 6: ecuaciones de estimación generalizadas sobre postests apilados y estandarizados con el pretest, agrupadas por aula; el efecto del programa sobre la comprensión literal en el mismo modelo fue de 0,111 DE.

15.4.8 Tamaños del efecto y logro del nivel satisfactorio

Los efectos estandarizados fueron moderados para el total inferencial (d_ppc2 = 0,43; g ajustada = 0,58) y para la inferencia de puente causal (0,48 y 0,58), menores para la elaborativo-evaluativa (0,34 y 0,43) y la léxica (0,33 y 0,36), y cercanos a cero para la comprensión literal (0,05 y 0,11), el

único resultado cuyos intervalos bootstrap incluyeron el cero (Tabla 15-9; Figura 15-2).

Tabla 15-9

Tamaños del efecto estandarizados con intervalos de confianza bootstrap por conglomerados

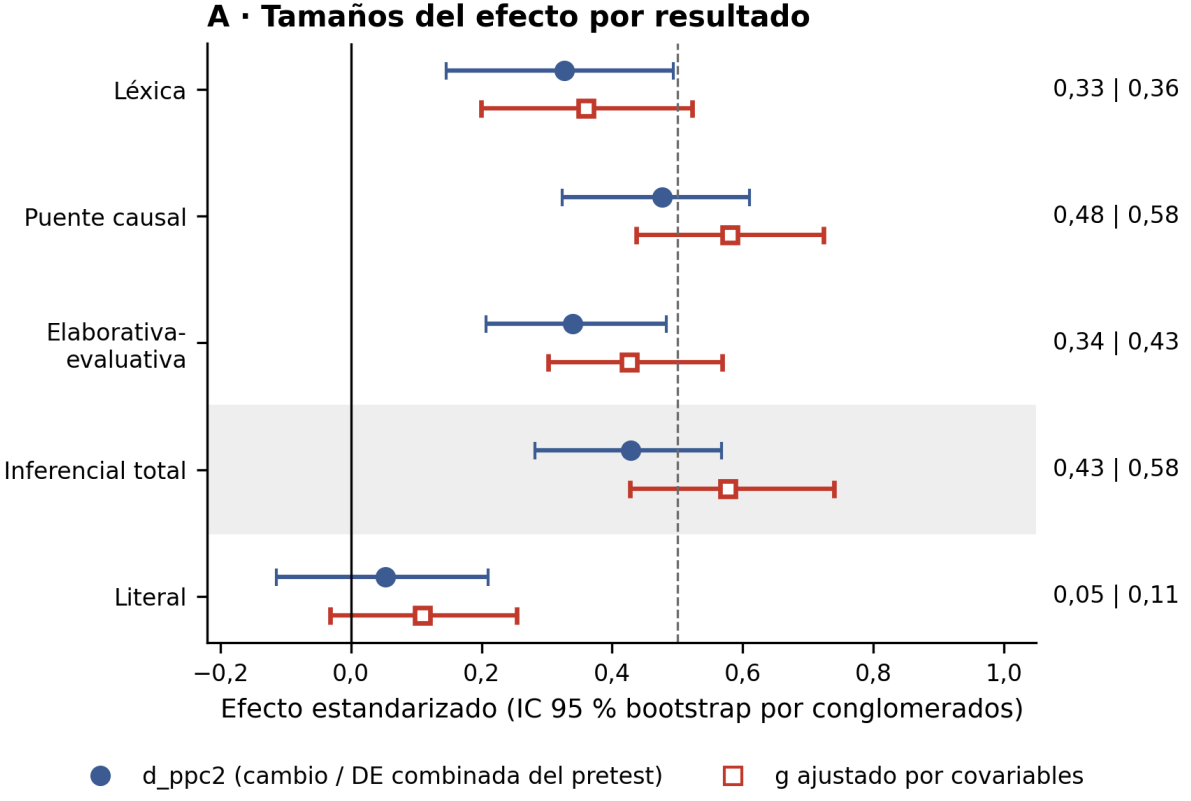
Resultado	d_ppc2	IC 95 % (bootstrap por aulas)	g ajustada	IC 95 % (bootstrap por aulas)
Inferencia léxica	0,327	[0,146; 0,494]	0,360	[0,199; 0,523]
Inferencia de puente causal	0,477	[0,324; 0,611]	0,581	[0,437; 0,724]
Inferencia elaborativo-evaluativa	0,340	[0,207; 0,483]	0,426	[0,302; 0,569]
Comprensión inferencial (total)	0,429	[0,282; 0,568]	0,577	[0,427; 0,740]
Comprensión literal	0,053	[–0,114; 0,210]	0,109	[–0,032; 0,254]

Nota: d_ppc2: diferencia entre los cambios medios de ambos brazos dividida por la DE combinada del pretest, con corrección de muestra pequeña (Morris, 2008). g ajustada: diferencia ajustada por covariables (pretest, grado, sexo, lengua dominante) dividida por la DE combinada del postest. Intervalos de 2.000 remuestreos bootstrap de aulas dentro de cada brazo (método percentil).

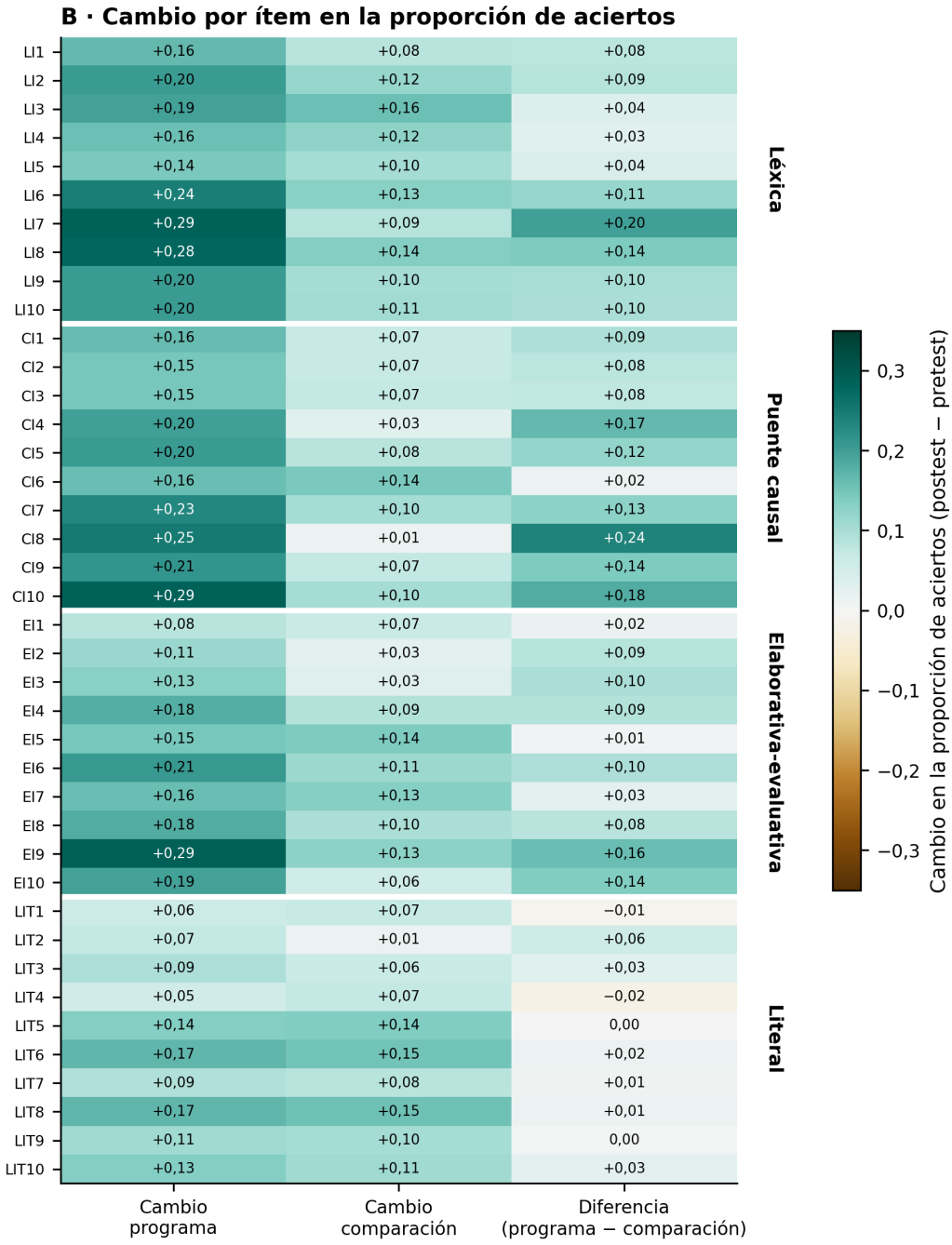
En el pretest alcanzaron el nivel satisfactorio (al menos 21 de 30 puntos inferenciales) el 30,7 % del alumnado del programa y el 29,9 % del de comparación; en el postest lo hicieron el 67,5 % y el 41,1 %, y la razón de momios ajustada de alcanzarlo fue 5,25 veces mayor en el programa (IC 95 % [3,25; 8,50]; $p < 0,001$). Entre quienes partieron de la banda básica, el 71 % del programa y el 35 % de la comparación llegaron a la banda satisfactoria (Figura 15-3).

Figura 15-2

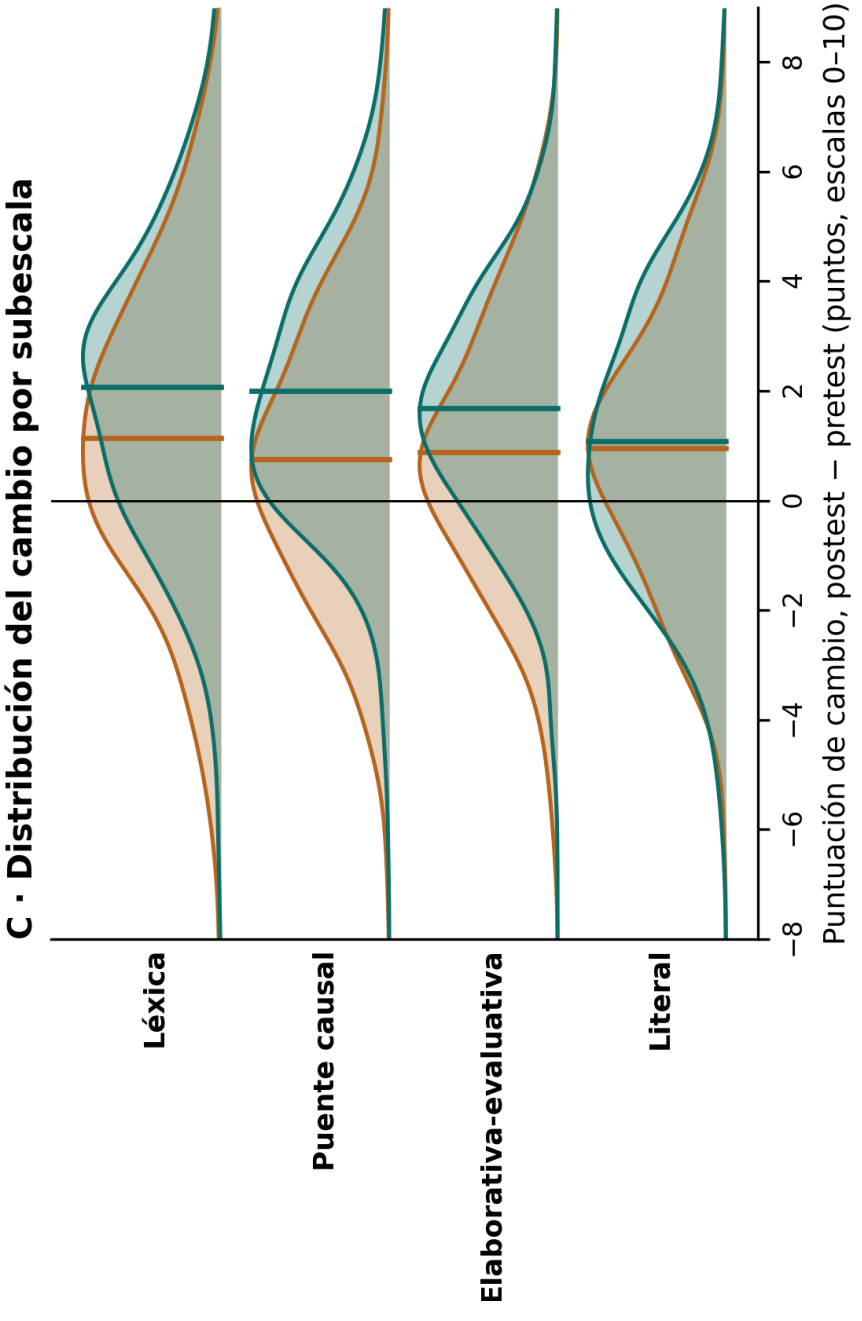
Efectos del programa por tipo de inferencia y por ítem (N = 990; 62 aulas)



Nota. Panel A: efectos estandarizados con IC 95 % bootstrap por conglomerados (2000 remuestreos): d_ppc2, diferencia del cambio medio dividida por la DE combinada del pretest (círculos), y g ajustado por covariables (cuadrados) para la inferencia léxica (0,33; 0,36), de puente causal (0,48; 0,58) y elaborativa-evaluativa (0,34; 0,43), la comprensión inferencial total (0,43; 0,58) y la comprensión literal (0,05; 0,11); línea discontinua en el efecto de planificación de 0,50.



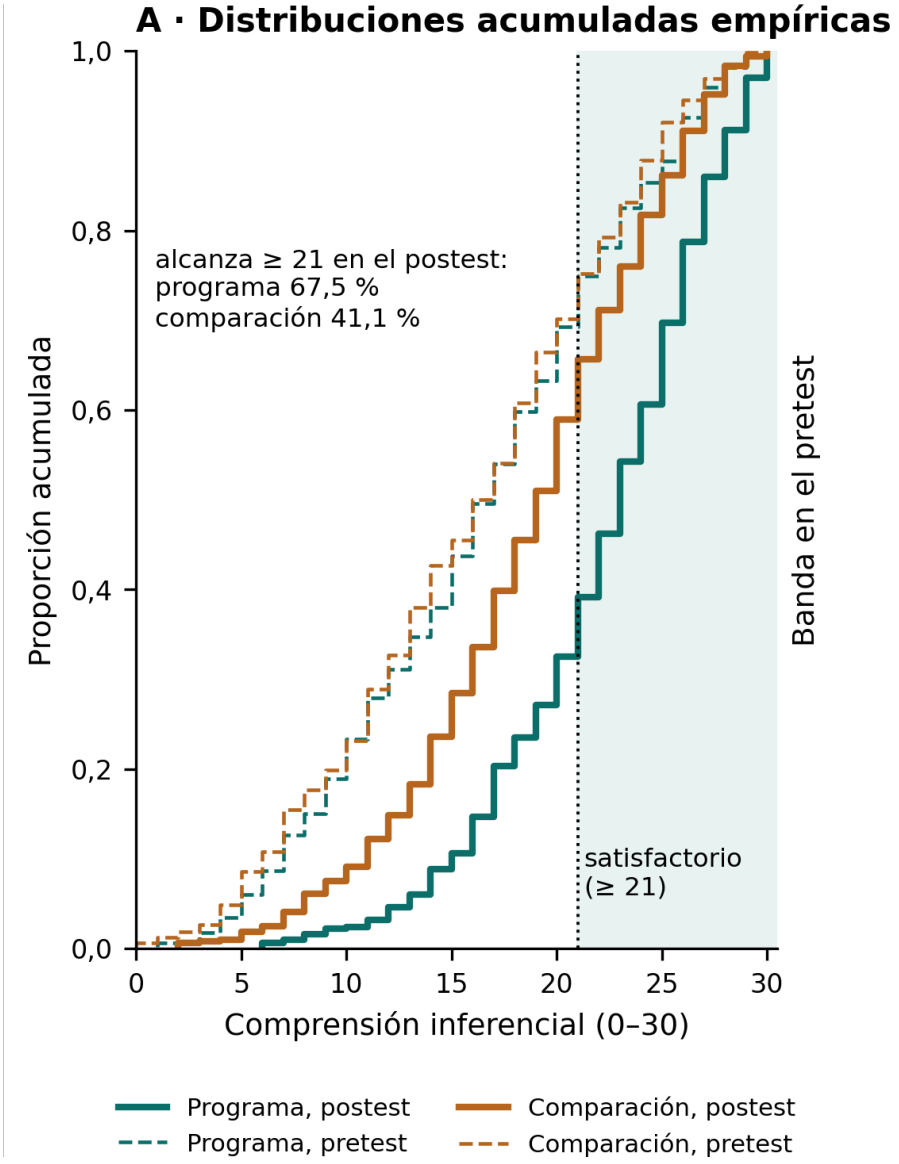
Nota. Panel B: cambio en la proporción de aciertos entre el pretest y el postest en cada uno de los 40 ítems, por brazo, y diferencia entre brazos; la ventaja del programa fue positiva en los 30 ítems inferenciales (de +0,01 a +0,24) y cercana a cero en los 10 ítems literales (de −0,02 a +0,06).



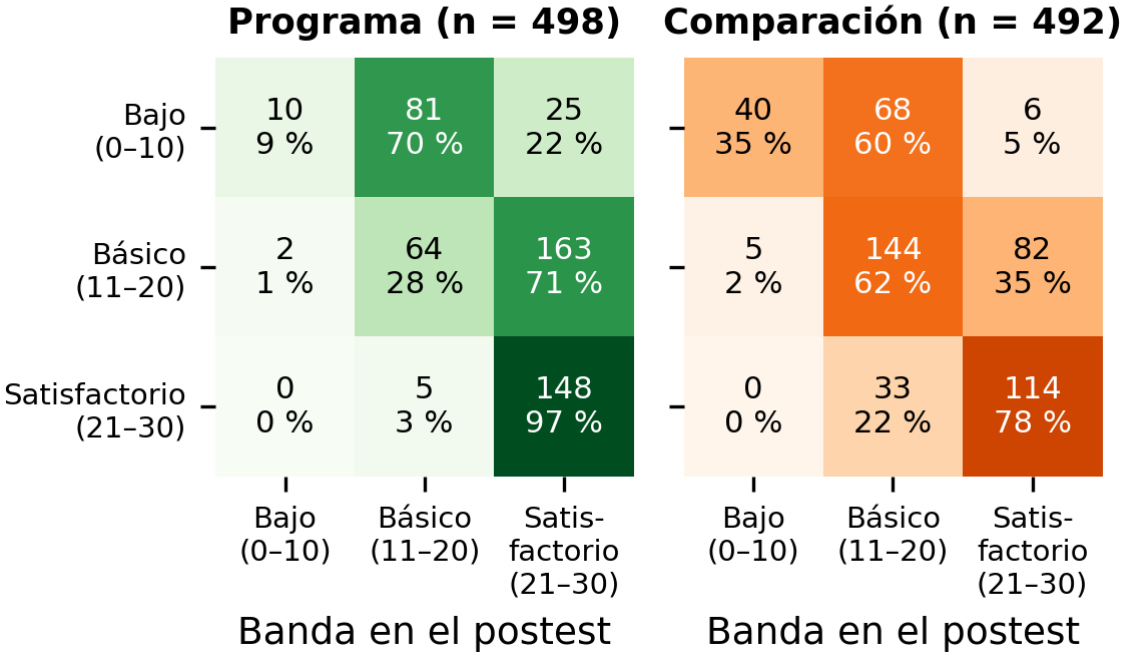
Nota. Panel C: densidades kernel de las puntuaciones individuales de cambio por subescala y brazo, con líneas verticales en las medias de cada brazo (programa frente a comparación: léxica 2,08 frente a 1,15; puente causal 2,00 frente a 0,76; elaborativa-evaluativa 1,68 frente a 0,88; literal 1,09 frente a 0,96).

Figura 15-3

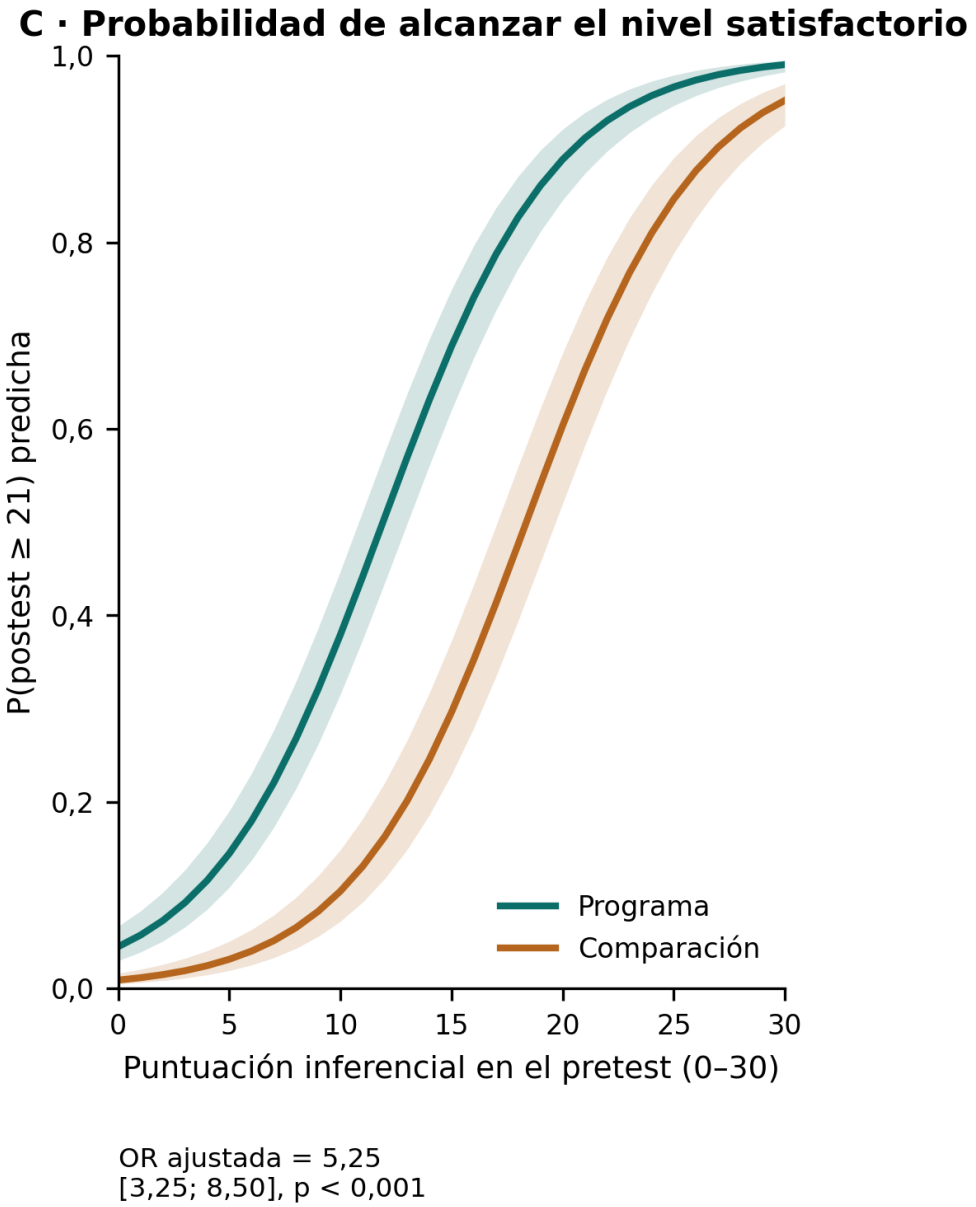
Distribución de las puntuaciones del postest y logro del nivel satisfactorio (N = 990; 62 aulas)



Nota. Panel A: distribuciones acumuladas empíricas de la comprensión inferencial en el pretest (discontinua) y en el postest (continua) por brazo; el área sombreada marca el nivel satisfactorio (al menos 21 de 30 puntos), que en el postest alcanzó el 67,5 % del alumnado del programa y el 41,1 % del de comparación.



Nota. Panel B: transiciones entre bandas de desempeño (bajo 0–10, básico 11–20, satisfactorio 21–30) del pretest (filas) al postest (columnas), con frecuencias y porcentajes por fila; entre quienes empezaron en la banda básica, el 71 % del programa y el 35 % de la comparación llegaron a la banda satisfactoria.



Nota. Panel C: probabilidad de alcanzar el nivel satisfactorio en el postest predicha por la ecuación de estimación generalizada logística con correlación intercambiable por aula, en función del pretest y con grado, sexo y lengua dominante en sus medias muestrales (bandas: IC 95 %); OR ajustada = 5,25; IC 95 % [3,25; 8,50]; p < 0,001.

15.4.9 Robustez del efecto primario y dosis-respuesta

El efecto del programa sobre la comprensión inferencial fue estable en todas las especificaciones, con estimaciones puntuales entre 3,135 y 3,174 puntos e intervalos de confianza del 95 % que excluyeron el cero en todas las que lo proporcionan (Tabla 15-10); la prueba de permutación aporta solo un valor p. El análisis a nivel de aula arrojó 3,17 puntos, la imputación múltiple de los 39 postests faltantes 3,17 puntos, y la prueba que reasignó las etiquetas de brazo entre las aulas dio p = 0,0005, el valor mínimo alcanzable con 2.000 permutaciones.

Tabla 15-10

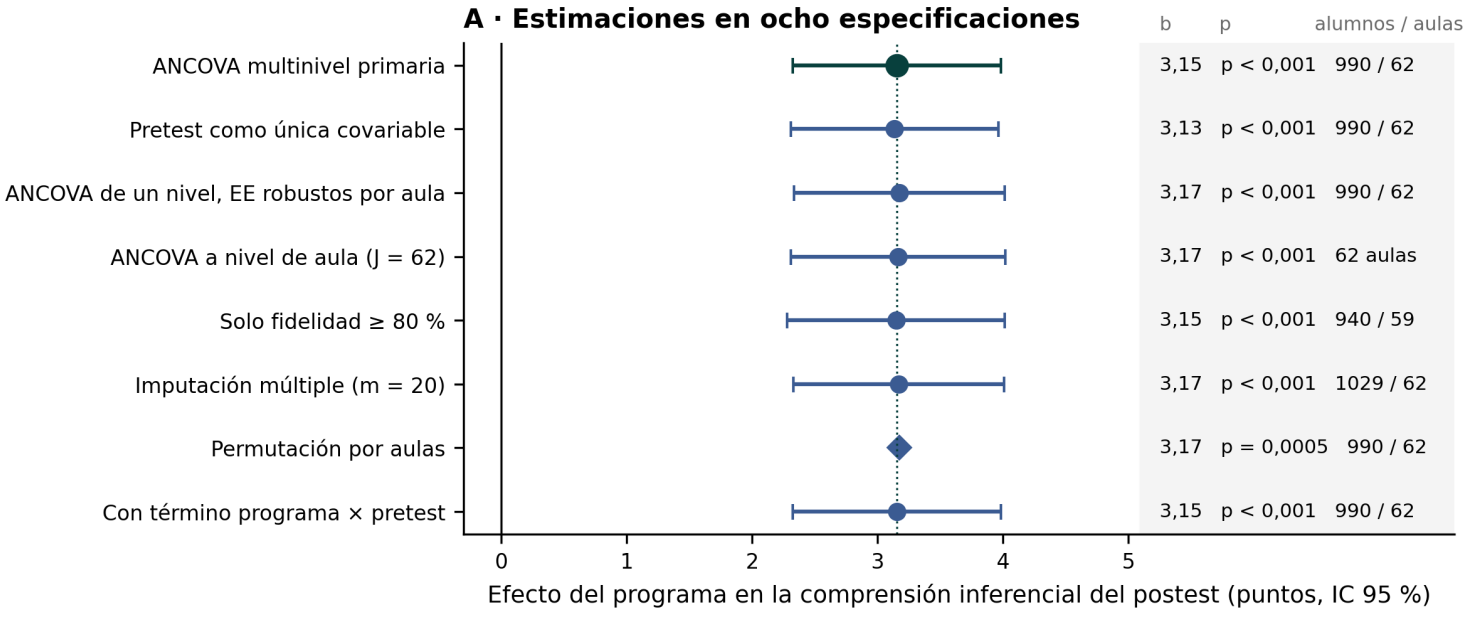
Robustez del efecto del programa sobre la comprensión inferencial del postest

Especificación	b	IC 95 %	p	Estudiantes / aulas
Primaria: ANCOVA multinivel (intercepto aleatorio por aula, REML)	3,154	[2,324; 3,984]	< 0,001	990 / 62
ANCOVA multinivel con el pretest como única covariable	3,135	[2,307; 3,962]	< 0,001	990 / 62
ANCOVA de un nivel con errores estándar robustos por conglomerado	3,174	[2,335; 4,014]	< 0,001	990 / 62
ANCOVA a nivel de aula sobre medias de aula, ponderada por tamaño	3,165	[2,309; 4,021]	< 0,001	62 aulas
Solo aulas del programa con fidelidad ≥ 80 %	3,148	[2,281; 4,014]	< 0,001	940 / 59
Imputación múltiple de postests faltantes (20 imputaciones, reglas de Rubin)	3,169	[2,331; 4,008]	< 0,001	1.029 / 62
Prueba de permutación por conglomerados (2.000 permutaciones de etiquetas de aula)	3,174	—	0,0005	990 / 62
ANCOVA multinivel con interacción programa × pretest (efecto en la media del pretest)	3,155	[2,326; 3,984]	< 0,001	990 / 62

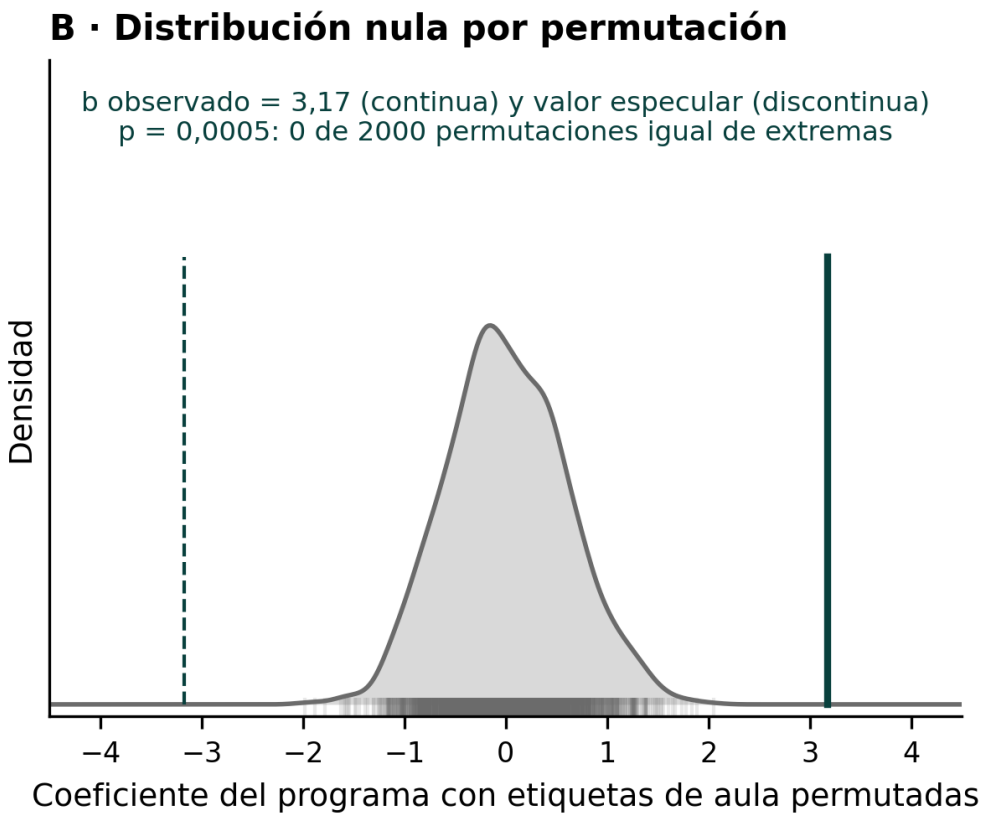
Nota: Todos los modelos se ajustan por el pretest centrado, grado, sexo y lengua dominante, salvo indicación. La fila de permutación informa la estimación de un nivel y el valor p de permutación. ANCOVA = análisis de covarianza; REML = máxima verosimilitud restringida..

Figura 15-4

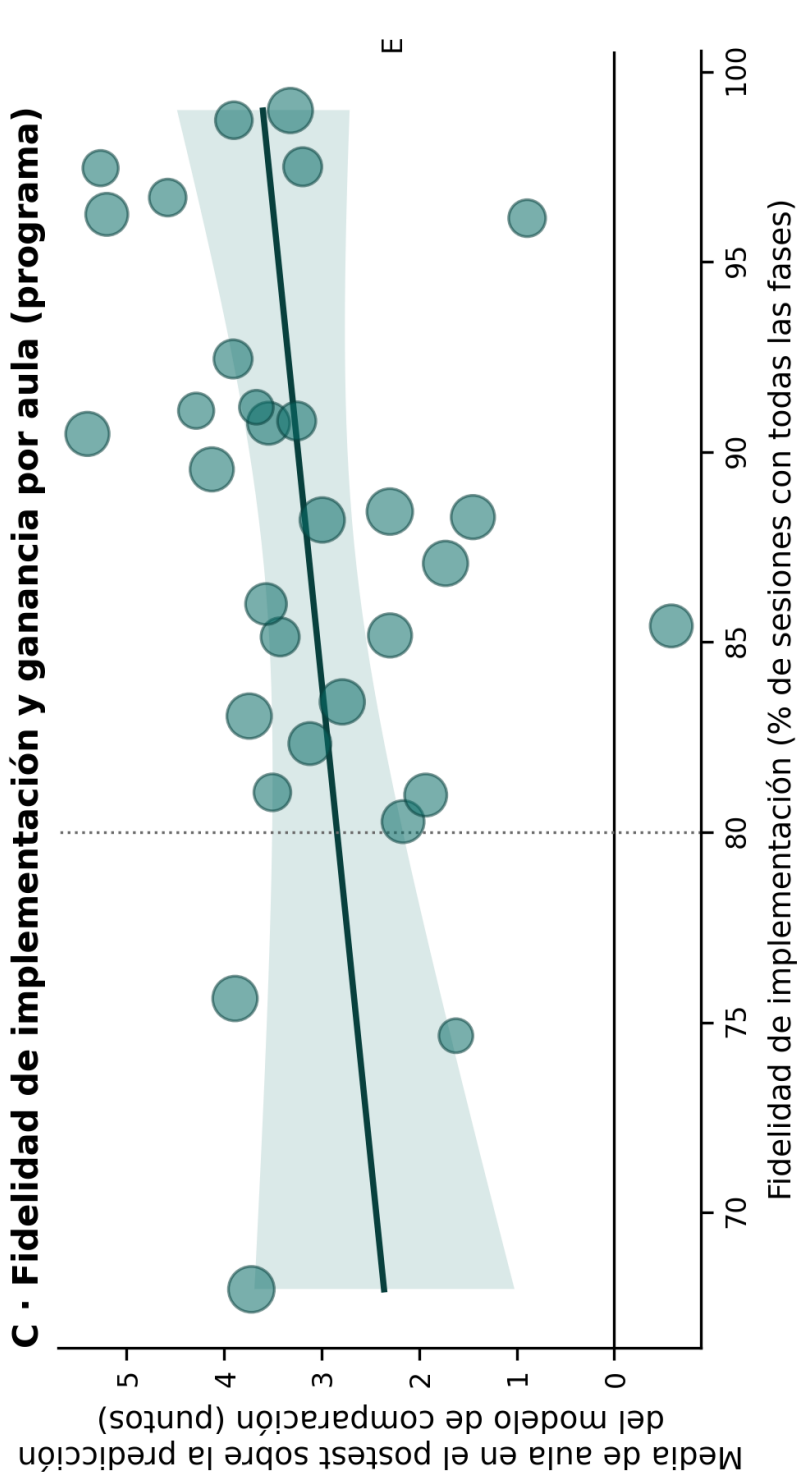
Robustez del efecto del programa, fidelidad de implementación y agrupamiento (N = 990; 62 aulas)



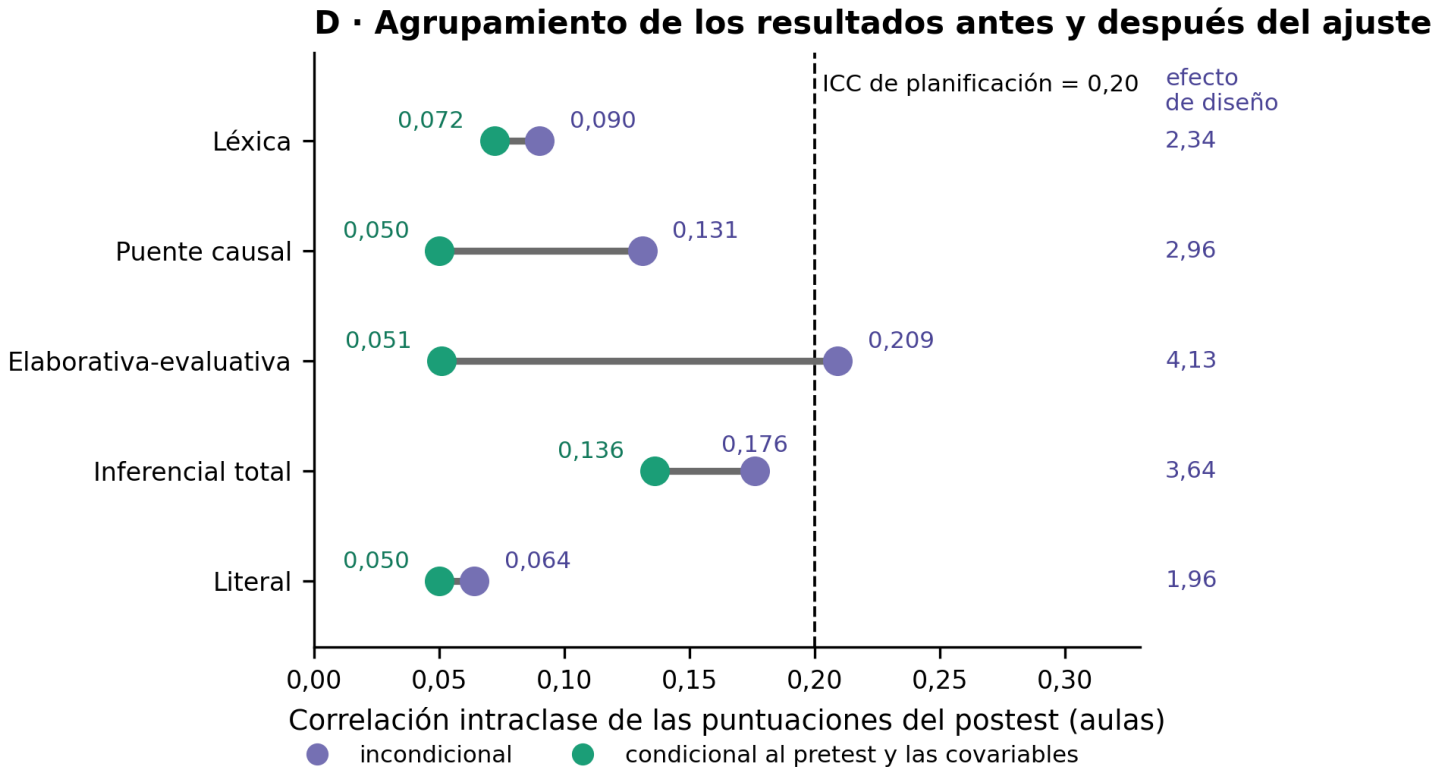
Nota. Panel A: efecto del programa en la comprensión inferencial del postest (puntos, IC 95 %) en el análisis de covarianza multinivel primario y en siete especificaciones alternativas (estimaciones de 3,13 a 3,17; la fila de permutación muestra la estimación de un nivel sin intervalo).



Nota. Panel B: distribución nula del coeficiente del programa en 2000 permutaciones de las etiquetas de brazo entre aulas (densidad kernel y marcas); el coeficiente observado (b = 3,17, línea continua) y su valor especular (línea discontinua) quedan más allá de todos los valores permutados, p = 0,0005.



Nota. Panel C: media del postest de cada aula del programa por encima de la predicción de un modelo multinivel ajustado en el brazo de comparación (pretest, grado, sexo y lengua dominante), según la fidelidad de implementación; área del marcador proporcional al tamaño del aula; línea y banda: ajuste por mínimos cuadrados ponderado por tamaño con IC 95 %; la estimación multinivel de dosis-respuesta a nivel de alumno fue de +0,42 puntos por cada 10 puntos porcentuales de fidelidad (IC 95 % [−0,19; 1,03]; p = 0,179; 498 alumnos, 31 aulas).



Nota. Panel D: correlaciones intraclase incondicionales y condicionales (pretest y covariables) de los cinco resultados del postest, con el efecto de diseño para el tamaño medio analítico de aula de 15,97 en la columna derecha; línea discontinua en el ICC de planificación de 0,20.

Al restringir el brazo del programa a las 28 aulas con fidelidad de al menos el 80 %, la estimación se mantuvo en 3,15 puntos. Dentro del brazo del programa, un aumento de 10 puntos porcentuales en la fidelidad se asoció con 0,42 puntos más en el postest (IC 95 % [–0,19; 1,03]; p = 0,179), una estimación de dosis-respuesta positiva pero imprecisa (Figura 15-4).

15.5 DISCUSIÓN

15.5.1 Una ganancia específica que no dependió de la lengua dominante

El hallazgo más desconcertante no es la magnitud del efecto sobre la inferencia, sino su especificidad. Un programa que moviliza el quechua y el saber comunitario podría haber elevado el rendimiento en toda la prueba, por mayor compromiso o por familiaridad con los contenidos; sin embargo, la comprensión literal cambió de manera similar en ambos brazos, y el contraste de especificidad situó el efecto inferencial 0,35 desviaciones estándar por encima del literal. Más inesperado aún, la diferencia bruta de ganancias entre brazos fue parecida en el alumnado quechua-dominante (3,02 puntos) y en el castellano-dominante (2,82 puntos), y la lengua dominante no se asoció con el postest una vez considerado el pretest. Estos patrones sugieren, de manera descriptiva, que recurrir al quechua no perjudicó a quienes dominan el castellano, y son consistentes con una lectura según la cual el programa actuó a través de los procesos que perseguía y no mediante un aumento general del desempeño en la prueba. La comparación por lengua dominante, no obstante, descansa en diferencias brutas de ganancia y no en una prueba formal de interacción, por lo que debe tomarse con cautela.

15.5.2 Magnitud del efecto y diálogo con la literatura previa

El efecto sobre la comprensión inferencial (d_ppc2 = 0,43; g ajustada = 0,58) es del mismo orden que las estimaciones metaanalíticas de la enseñanza de inferencias, que arrojaron d = 0,68 para la comprensión inferencial y d = 0,58 para la general (Elleman, 2017), aunque la estimación d_ppc2 queda algo por debajo de ese rango y solo la g ajustada lo alcanza en su

extremo inferior. Esa convergencia resulta llamativa si se considera que el programa lo condujeron docentes regulares de aulas rurales multigrado, en una segunda lengua para la mayoría del alumnado y en una sola etapa lectiva. En contraste, un programa reciente que combina lenguaje y lectura de palabras informó un efecto de SMD = 0,26 sobre la comprensión lectora (Nergård et al., 2026), y la intervención inferencial con tutores en secundaria produjo g = 0,40 en los tipos de inferencia enseñados (Barnes et al., 2024). Sería prematuro leer esa diferencia como superioridad del presente programa: la medida aquí empleada estuvo alineada con los procesos enseñados, las medidas amplias de comprensión tienden a diluir los efectos de la instrucción focalizada, y las poblaciones, las duraciones y los diseños difieren de manera sustancial. Tampoco los programas de conocimiento, lenguaje e indagación para estudiantes multilingües (Davis et al., 2025; Relyea et al., 2025), impartidos en pequeños grupos por docentes de inglés como segunda lengua, permiten ordenar magnitudes. Su valor para esta discusión es otro: muestran que construir conocimiento y lengua al servicio de la comprensión es factible con alumnado multilingüe, y el presente trabajo sugiere que una versión entregada a aulas completas por docentes regulares, y anclada en la lengua y el saber comunitario, también puede producir ganancias. Las comparaciones entre intervenciones mediadas por ordenador que contrastan la instrucción estratégica (Serrano et al., 2023) y la evidencia sobre la enseñanza culturalmente responsiva, donde la comprensión, el lenguaje y la decodificación responden a la enseñanza estructurada (Sinclair, 2025), son compatibles con un efecto de la enseñanza estructurada, pero no resuelven si el componente cultural añade algo a la rutina inferencial; el diseño actual tampoco permite aislarlo.

Desde una perspectiva crítica, el debate entre estrategia y conocimiento ilumina este punto. La reevaluación de la evidencia sobre la instrucción estratégica frente a los argumentos de que el conocimiento es la clave de la comprensión (Dewitz y Graves, 2024) y la importancia del conocimiento documentada en un estudio cuasiexperimental con alumnado de primaria (Smith et al., 2026) no se excluyen entre sí. Los resultados presentes ocupan una posición intermedia: el programa no enseñó estrategias en el vacío

ni transmitió solo conocimiento, sino que activó el saber comunitario antes de modelar la inferencia, y el patrón de ganancias inferenciales sin ganancias literales es consistente con que ambos componentes sean complementarios. Sin un brazo que enseñe inferencias sin conocimiento comunitario, esta lectura sigue siendo una hipótesis plausible y no una conclusión.

15.5.3 Efectos diferenciados entre tipos de inferencia

El efecto fue mayor en la inferencia de puente causal (g = 0,58) que en la elaborativo-evaluativa (0,43) y la léxica (0,36). Las inferencias de puente conectan sucesos y estados que el texto presenta por separado, y el pensamiento en voz alta y los organizadores de causa y efecto del programa hacen explícita esa conexión; además, el conocimiento requerido para vincular causas y consecuencias en textos sobre calendarios agrícolas, trabajo comunal o narrativas locales es precisamente el que la fase de discusión en quechua trae a la lección (Rice et al., 2024; Smith et al., 2021). Las inferencias elaborativo-evaluativas dependen de un conocimiento más amplio y de la capacidad de juicio, que probablemente se desarrollan con mayor lentitud y resultan más difíciles de modificar en una sola etapa lectiva. En el caso de la inferencia léxica, el vocabulario temprano de los niños bilingües predice la comprensión posterior dentro de cada lengua y no entre lenguas (Giguere y Hoff, 2024), de modo que catorce semanas pueden resultar insuficientes para que el alumnado acumule la exposición al vocabulario castellano que esa inferencia exige. La transferencia translingüística de las habilidades de palabra, por su parte, sigue siendo objeto de debate (Xie y Yeung, 2026), y una prueba aplicada solo en castellano no permite adjudicar si las ganancias inferenciales se trasladarían a la lengua materna. Que los tres efectos fueran significativos tras la corrección mientras la comprensión literal no cambiaba diferencialmente apoya la idea de que el programa operó sobre los procesos enseñados, aunque este argumento no descarta otros mecanismos no medidos.

15.4.4 Translenguaje y pedagogías culturalmente sostenibles

Estudios cualitativos y exploratorios describen el translenguaje como un medio para legitimar el conocimiento minorizado en el aula: en escuelas etnoeducativas de Colombia (Quinn et al., 2026), en la práctica de docentes indígenas que incorporan su lengua a lecciones en la lengua mayoritaria en Vietnam (Nguyen, 2025) y en las perspectivas de educadores sobre alumnado indígena y refugiado en Brasil (Angelo et al., 2026). Esos trabajos son valiosos por su densidad descriptiva, pero rara vez miden resultados de comprensión; el presente estudio aporta una estimación cuantitativa que es compatible con la tesis de que translenguaje y pedagogías culturalmente sostenibles dependen mutuamente (Beiler y Villacañas, 2025). Conviene, con todo, no sobrerrepresentar esa convergencia. La intervención combinó la enseñanza explícita de inferencias, la activación del conocimiento comunitario y la discusión en quechua, y el diseño no permite atribuir la ganancia a un componente en particular; la lectura más prudente es que el paquete funciona y que la contribución separada del translenguaje y de la dimensión cultural queda por determinar.

15.5.5 Equidad e implicaciones para la educación intercultural bilingüe

El efecto sobre la comprensión inferencial no dependió del nivel inicial: fue de unos tres puntos tanto a una desviación estándar por debajo como por encima de la media y resultó significativo en todo el rango del pretest. Solo en la inferencia de puente causal el efecto fue algo mayor para quienes partían de niveles más bajos, patrón que probablemente refleja en parte el techo de la subescala en el brazo del programa. En un contexto con niveles iniciales bajos, donde menos de uno de cada tres estudiantes alcanzaba el nivel satisfactorio, la ventaja de más de 25 puntos porcentuales en el postest tiene relevancia para la equidad. Estos resultados son consistentes con la evidencia de que la instrucción estructurada de la comprensión beneficia a lectores de distintos niveles (Peng et al., 2024; Vaughn y Barnes, 2023).

Para la EIB, los hallazgos ofrecen evidencia basada en el diseño de que sus principios pedagógicos, la instrucción que se apoya en la primera lengua

y en los saberes locales, pueden traducirse en una rutina estructurada y replicable con ganancias medibles en comprensión de orden superior. Ello interesa a la reforma en curso de la formación docente para la EIB en el Perú (Chávez et al., 2026; Moreano et al., 2026) y a otros sistemas donde la infancia indígena aprende a leer en una lengua dominante (Limerick, 2023). De manera más general, el estudio respalda una concepción de la enseñanza de la comprensión para aprendices multilingües en la que la lengua del hogar no es un andamio que deba retirarse, sino un canal por el que el conocimiento necesario para inferir llega al texto (Budde et al., 2026; Davis et al., 2025). Los componentes del programa (pensamiento en voz alta bilingüe, activación del conocimiento en la lengua del hogar y organizadores gráficos bilingües) no dependen del par de lenguas concreto, y su adaptación a otros contextos bilingües indígenas e inmigrantes constituye una vía razonable de transferencia, aunque no probada aquí.

15.5.6 Limitaciones y líneas futuras

Cuatro limitaciones matizan las conclusiones:

1) La asignación al programa no fue aleatoria. Los brazos fueron equivalentes o casi equivalentes en la línea base, las estimaciones se ajustaron por el pretest y las covariables, y el efecto resistió especificaciones que incluyen los análisis de permutación y de aula, pero no pueden descartarse diferencias no observadas entre las escuelas que se incorporaron al programa y las de comparación; la réplica con aleatorización por conglomerados es el paso natural.

2) Se usó el mismo formulario en ambas olas: los efectos de práctica afectarían a los dos brazos, pero pueden haber contribuido a las ganancias de la comparación, y la compresión de las puntuaciones cerca del techo en algunas subescalas del brazo del programa pudo atenuar los efectos estimados.

3) La comprensión se midió solo en castellano; una medida paralela en quechua mostraría si las ganancias inferenciales aparecen también en la lengua del hogar y si el programa favorece la alfabetización bilingüe.

4) El postest se aplicó al término de la intervención, por lo que se desconoce la persistencia de los efectos, y la estimación de dosis-respuesta fue positiva pero imprecisa. Los estudios futuros deberían seguir a la cohorte en la educación secundaria, examinar mediadores docentes de la fidelidad y someter a prueba el programa con otras lenguas indígenas.

15.6 CONCLUSIONES

Un programa intercultural bilingüe de lectura inferencial de 14 semanas mejoró la comprensión inferencial del alumnado quechua–castellano de quinto y sexto grado en aproximadamente media desviación estándar, con efectos sobre la inferencia léxica, la de puente causal y la elaborativo-evaluativa, sin cambio diferencial en la comprensión literal y con una ventaja de 26 puntos porcentuales en la proporción de estudiantes que alcanzaron el nivel satisfactorio. El efecto fue estable en los análisis multinivel, de aula, de imputación y de permutación, y a lo largo del rango de habilidad inicial, lo que sugiere que la mejora no fue un artefacto de la especificación analítica ni se concentró en un tramo del alumnado.

La especificidad del patrón apunta a que el programa actuó sobre los procesos inferenciales que enseñaba, en la línea de la hipótesis de que la lengua materna y el saber comunitario funcionan como conocimiento disponible para construir el modelo de la situación. Anclar la enseñanza de inferencias en el repertorio lingüístico y cultural del alumnado se perfila, así, como una estrategia prometedora y escalable para elevar la comprensión de orden superior en la educación intercultural bilingüe. Para la formación y el acompañamiento docente, el resultado señala el valor de rutinas estructuradas y replicables, como el pensamiento en voz alta bilingüe, la activación del conocimiento comunitario y los organizadores gráficos bilingües, que traducen los principios de la EIB en práctica de aula.

Estas conclusiones se circunscriben a las aulas analizadas, dado el carácter no aleatorio de la asignación, y requieren réplicas con aleatorización por conglomerados, seguimiento longitudinal y medidas en quechua. Con

esas salvedades, los hallazgos respaldan, más allá del caso andino, tratar los repertorios lingüísticos y culturales de la infancia como recursos cognitivos para la comprensión en los sistemas educativos donde se aprende a leer en una lengua distinta de la propia.

15.7 REFERENCIAS BIBLIOGRÁFICAS

Angelo, M., Oliveira, K., Regina, C., Braunstein, L., & Donley, K. (2026). 'Essa é a nossa luta!' Educators' perspectives on translanguaging as an inclusive pedagogy for Indigenous and refugee multilingual learners in Brazilian public schools. *The Language Learning Journal*, 1-18. https://doi.org/10.1080/09571736.2026.2695466

Barnes, M., Clemens, N., Simmons, D., Hall, C., Fogarty, M., Martinez, A., Vaughn, S., Simmons, L., Fall, A., & Roberts, G. (2024). A Randomized Controlled Trial of Tutor- and Computer-Delivered Inferential Comprehension Interventions for Middle School Students with Reading Difficulties. *Scientific Studies of Reading*, *28*(4), 411-440. https://doi.org/10.1080/10888438.2024.2331517

Beiler, I., & Villacañas, L. (2025). Translanguaging and culturally sustaining pedagogies: A mutually dependent relationship?. *Foreign Language Annals*, *58*(1), 159-181. https://doi.org/10.1111/flan.12791

Borm, G., Fransen, J., & Lemmens, W. (2007). A simple sample size formula for analysis of covariance in randomized clinical trials. *Journal of Clinical Epidemiology*, *60*(12), 1234-1238. https://doi.org/10.1016/j.jclinepi.2007.02.006

Budde, C., Amendum, S., & Kittle, J. (2026). An Enhanced Science of Reading for Multilingual Learners: The Importance of Leveraging

Multilingualism. *The Reading Teacher*, *79*(5). https://doi.org/10.1002/trtr.70041

Chávez, R., Ruiz, Y., Carranza, R., Campos, J., Guevara, R., Navas, C., & Poémape, C. (2026). Intercultural bilingual education in teacher education within native communities of the Amazonas region intercultural bilingual education in teacher education within native communities of the Amazonas region. *Studies in Higher Education*, 1-20. https://doi.org/10.1080/03075079.2026.2679171

Davis, D., Kelcey, B., Dobis, C., Huang, B., & Relyea, J. (2025). Evaluating the Impact of the Knowledge, Language, and Structured Inquiry (K.L.I.) Intervention on Reading Comprehension in Upper Elementary Multilingual English Learners. *The Journal of Experimental Education*, 1-27. https://doi.org/10.1080/00220973.2025.2590449

Dewitz, P., & Graves, M. (2024). Reevaluating and Restructuring Comprehension Strategy Instruction. *The Reading Teacher*, *78*(1), 17-26. https://doi.org/10.1002/trtr.2308

Duke, N., Ward, A., & Pearson, P. (2021). The Science of Reading Comprehension Instruction. *The Reading Teacher*, *74*(6), 663-672. https://doi.org/10.1002/trtr.1993

Elleman, A. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology*, *109*(6), 761-781. https://doi.org/10.1037/edu0000180

Giguere, D., & Hoff, E. (2024). Bilingual children's vocabulary skills at 5 years predict reading comprehension development within, not across, languages. *International Journal of Bilingual Education and Bilingualism, 27*(2), 240-252. https://doi.org/10.1080/13670050.2023.2166397

Guevara, R., Diaz, M., Tongo, D., Bautista, W., Quiac, S., & De La Puente, E. (2026). Early literacy profiles of Awajún children in the Peruvian Amazon: evidence from intercultural bilingual schools. *Reading and Writing*. https://doi.org/10.1007/s11145-026-10885-3

Hedges, L., & Hedberg, E. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87. https://doi.org/10.3102/0162373707299706

Limerick, N. (2023). Linguistic Registers and Citizenship Education: Divergent Approaches to Content, Instruction, Kichwa Use, and State Relationships in Ecuador's Intercultural Bilingual Education. *American Educational Research Journal, 60*(2), 219-256. https://doi.org/10.3102/00028312231152584

Liñán, H., Sano, J., Paucar, C., & Tafur, L. (2023). Intercultural Bilingual Education (EIB) in Peru. *Research in Subject-matter Teaching and Learning (RISTAL), 6*(1), 84-95. https://doi.org/10.2478/ristal-2023-0006

McDonald, A., Morrison, T., Wilcox, B., & Billen, M. (2021). Improving Children's Reading Comprehension by Teaching Inferences. *Reading Psychology, 42*(3), 264-280. https://doi.org/10.1080/02702711.2021.1888351

Moreano, G., Zegarra, C., & Rivera, Y. (2026). New intercultural bilingual education teacher preparation curriculum: Opportunities and challenges to reduce educational inequities in Peru. *Education Policy Analysis Archives, 34*. https://doi.org/10.14507/epaa.34.9008

Morris, S. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods, 11*(2), 364-386. https://doi.org/10.1177/1094428106291059

Nergård, T., Furnes, B., Caglar, Ø., Friborg, O., & Melby, M. (2026). Improving reading comprehension through language comprehension and early word reading: A multisite randomized trial. *Journal of Educational Psychology, 118*(4), 484-503. https://doi.org/10.1037/edu0001017

Nguyen, T. (2025). Voicing decolonial dialogues: Indigenous teachers' translanguaging in the mainstream classroom. *Linguistics and Education, 88,* 101440. https://doi.org/10.1016/j.linged.2025.101440

O'Hare, L., Stark, P., Cockerill, M., Lloyd, K., McConnellogue, S., Gildea, A., Biggart, A., Bower, C., & Connolly, P. (2023). Comparing the effectiveness of two reciprocal reading comprehension interventions for primary school pupils in disadvantaged schools. *British Journal of Educational Psychology, 93*(4), 1123-1145. https://doi.org/10.1111/bjep.12623

Peng, P., Wang, W., Filderman, M., Zhang, W., & Lin, L. (2024). The Active Ingredient in Reading Comprehension Strategy Intervention for Struggling Readers: A Bayesian Network Meta-analysis. *Review of Educational Research, 94*(2), 228-267. https://doi.org/10.3102/00346543231171345

Quinn, P., Vinueza, C., & Justin, N. (2026). Exploring Pedagogical Translanguaging in EFL: An Exploratory Study in Colombian Ethno-Educational Schools. *International Journal of Applied Linguistics, 36*(2), 1782-1793. https://doi.org/10.1111/ijal.70032

Qureshi, M., & Aljanadbah, A. (2022). Translanguaging and reading comprehension in a second language. *International Multilingual Research Journal*, *16*(4), 247-257. https://doi.org/10.1080/19313152.2021.2009158

Relyea, J., Davis, D., Dobis, C., Huang, B., & Samuelson, C. (2025). Feasibility of the Knowledge, Language, and Inquiry (K.L.I.) intervention for multilingual English learners. *The Journal of Educational Research*, *118*(2), 131-146. https://doi.org/10.1080/00220671.2024.2449035

Rice, M., Wijekumar, K., Lambright, K., & Bristow, A. (2024). Inferencing in Reading Comprehension: Examining Variations in Definition, Instruction, and Assessment. *Technology, Knowledge and Learning*, *29*(3), 1169-1190. https://doi.org/10.1007/s10758-023-09660-y

Serrano, M., Villalón, R., Melero, Á., & Izquierdo, B. (2023). Effects of two computer-based interventions on reading comprehension: Does strategy instruction matter?. *Computers & Education*, *196*, 104727. https://doi.org/10.1016/j.compedu.2023.104727

Sinclair, J. (2025). Exploring the Effects of Culturally Responsive Instruction on Reading Comprehension, Language Comprehension, and Decoding with Bayesian Multilevel Models. *Education Sciences*, *15*(11), 1560. https://doi.org/10.3390/educsci15111560

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The Role of Background Knowledge in Reading Comprehension: A Critical Review. *Reading Psychology*, *42*(3), 214-240. https://doi.org/10.1080/02702711.2021.1888348

Smith, R., Snow, P., Serry, T., & Hammond, L. (2026). The Importance of Knowledge for Reading Comprehension: A Quasi-Experimental Investigation. *Reading Psychology*, *47*(7), 1111-1137. https://doi.org/10.1080/02702711.2026.2644971

Van Zeijts, B., Ganushchak, L., De Koning, B., & Tabbers, H. (2024). Stimulating inference-making in second grade children when reading and listening to narrative texts. *Reading and Writing*, *37*(9), 2201-2227. https://doi.org/10.1007/s11145-023-10463-x

Vaughn, S., & Barnes, M. (2023). Reading comprehension for students with reading disabilities: Progress and challenges. *Learning and Individual Differences*, *102*, 102258. https://doi.org/10.1016/j.lindif.2023.102258

Wawire, B., Barnes, A., Liang, X., & Piper, B. (2024). Supporting multilingual children at-risk of reading failure: impacts of a multilingual structured pedagogy literacy intervention in Kenya. *Reading and Writing*, *37*(8), 1975-2005. https://doi.org/10.1007/s11145-023-10453-z

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Handbooks

Xie, Q., & Yeung, S. (2026). Do word-level skills show cross-linguistic transfer and predict bilingual reading?. *Reading and Writing*, *39*(8), 3275-3299. https://doi.org/10.1007/s11145-026-10767-8

CAPÍTULO XVI

EDUCACIÓN INTERCULTURAL BILINGÜE Y COMPRENSIÓN LECTORA: ESTUDIO LONGITUDINAL EN COMUNIDADES ESCOLARES RURALES ANDINAS DEL PERÚ, 2024-2026

Lina Lucy Fabian Ojeda

Universidad Nacional de Educación Enrique Guzmán y Valle

Karla Sofia Dongo Ortiz

Universidad Nacional de Educación Enrique Guzmán y Valle

Fiorella Ñahuin Antonio

Universidad Nacional de Educación Enrique Guzmán y Valle

Resumen

La heterogeneidad con que se implementa la educación intercultural bilingüe (EIB) y la incertidumbre sobre si la alfabetización en lengua originaria favorece o resta a la del castellano plantean una pregunta a la vez de política y de teoría. Este estudio estimó cómo la intensidad de implementación escolar de la EIB se asocia con el nivel inicial y el crecimiento de la comprensión lectora en ambas lenguas, y separó, dentro de cada estudiante, la transferencia entre lenguas de las diferencias estables. Se siguió un panel de tres olas anuales (2024-2026) con 2 565 estudiantes quechuahablantes y aimarahablantes de 219 escuelas rurales de Apurímac, Ayacucho, Cusco y Puno, de tercero a quinto grado. Se contrastó la invarianza longitudinal, se estimó un modelo de crecimiento latente en paralelo con un índice escolar de implementación y un modelo de panel con interceptos aleatorios, con información completa e inferencia por bootstrap de escuelas. La medida fue invariante a nivel escalar en ambas lenguas. Una desviación estándar más de implementación se asoció con un nivel inicial mayor en lengua originaria (b = 0,64) y con un crecimiento anual mayor en castellano (b = 0,44); el nivel inicial en lengua originaria predijo ese crecimiento (β = 0,42) y transmitió parte del efecto de la implementación. Dentro de cada estudiante, el progreso en lengua originaria precedió al progreso en castellano (β = 0,24–0,26) más que a la inversa (β = 0,11–0,14). Los resultados sugieren que la alfabetización en lengua originaria no compite con la del castellano, sino que la antecede, y que la calidad de la implementación es una palanca escolar asociada con ambas trayectorias.

Palabras clave: educación intercultural bilingüe; comprensión lectora; interdependencia lingüística; modelo de crecimiento latente; modelo de panel con interceptos aleatorios; quechua; aimara.

Abstract

The uneven implementation of intercultural bilingual education (IBE) and the uncertainty over whether literacy in an Indigenous language supports or detracts from literacy in Spanish raise a question of both policy and theory. This study estimated how the intensity of school-level IBE implementation is associated with the initial level and growth of reading comprehension in both languages, and separated, within students, cross-language transfer from stable differences between students. A three-wave annual panel (2024-2026) followed 2,565 Quechua- and Aymara-speaking students from 219 rural schools in Apurímac, Ayacucho, Cusco, and Puno, from third to fifth grade. Longitudinal measurement invariance was tested, and a parallel latent growth model with a school-level implementation index and a random intercept cross-lagged panel model were estimated, using full-information estimation and school-level bootstrap inference. The measure was scalar invariant in both languages. One standard deviation more implementation was associated with a higher initial level in the Indigenous language ($b = 0.64$) and with greater annual growth in Spanish ($b = 0.44$); the initial Indigenous-language level predicted that growth ($\beta = 0.42$) and transmitted part of the implementation effect. Within students, progress in the Indigenous language preceded progress in Spanish ($\beta = 0.24–0.26$) more than the reverse ($\beta = 0.11–0.14$). The results suggest that literacy in the Indigenous language does not compete with literacy in Spanish but precedes it, and that implementation quality is a school-level lever associated with both trajectories.

Keywords: *intercultural bilingual education; reading comprehension; linguistic interdependence; latent growth model; random intercept cross-lagged panel model; Quechua; Aymara.*

16.1 INTRODUCCIÓN

16.1.2 Planteamiento del problema

Leer con comprensión al terminar la primaria es una condición para el aprendizaje posterior en todas las áreas, y la brecha en esta competencia entre estudiantes que hablan una lengua indígena y estudiantes que

crecen en la lengua dominante es una de las desigualdades educativas más persistentes de América Latina. En el Perú, la política de educación intercultural bilingüe (EIB) organiza la enseñanza de la lectura primero en la lengua originaria y después en castellano como segunda lengua, y se aplica sobre todo en escuelas rurales, donde los aprendizajes dependen además de factores socioeconómicos y escolares con una marcada distribución territorial (Flores et al., 2023; Paredes et al., 2026). La política existe, pero su implementación es heterogénea entre escuelas: varían el tiempo de instrucción en lengua originaria, la disponibilidad de materiales y el dominio de la lengua por parte de los docentes (Kvietok, 2023; Moreano et al., 2026; Tarrillo et al., 2026), y en la región la EIB oscila entre la asimilación y el reconocimiento de los pueblos indígenas (Zsögön, 2025). Esa heterogeneidad convierte una pregunta de política en una pregunta empírica precisa: si una implementación más intensa se traduce en mejores trayectorias de lectura, y en cuál de las dos lenguas.

La pregunta tiene una contraparte teórica de alcance general. La hipótesis de la interdependencia lingüística sostiene que las competencias académicas desarrolladas en una lengua se apoyan en una base cognitiva compartida y se transfieren a la otra (Cummins, 1979). Si es así, invertir en la alfabetización en lengua originaria no resta tiempo a la del castellano: la prepara. Si no lo es, los dos aprendizajes compiten por un tiempo escolar escaso y la enseñanza en lengua originaria tendría un costo de oportunidad. La evidencia disponible no permite zanjar la cuestión por tres motivos que este estudio aborda directamente.

16.1.3 Antecedentes y vacío de conocimiento

1) La mayor parte de la evidencia sobre transferencia es correlacional y transversal. La síntesis de Melby y Lervåg (2011) reunió la evidencia correlacional entre lenguas con gran heterogeneidad, lo que es compatible tanto con transferencia como con diferencias estables entre estudiantes que afectan a las dos lenguas a la vez. Los estudios longitudinales recientes describen trayectorias de comprensión en primera y segunda lengua desde la educación inicial hasta tercer grado

(Giguere y Hoff, 2024; Kwakkel et al., 2024) y la cambiante contribución de las habilidades de cada lengua con el tiempo (Shen y Huang, 2026; Xie y Yeung, 2026), pero casi siempre con pares de lenguas indoeuropeas o asiáticas y en contextos escolarizados de altos recursos.

2) Cuando la pregunta es si el progreso en una lengua antecede al progreso en la otra, el modelo de panel con efectos cruzados tradicional confunde dos fuentes de covariación: las diferencias estables entre estudiantes y la dinámica dentro de cada estudiante a lo largo del tiempo (Hamaker et al., 2015). Los efectos cruzados estimados sin separar esas fuentes pueden sobrestimar o invertir la transferencia, y la separación intrapersona e interpersona, mediante interceptos aleatorios o curvas latentes con residuos estructurados, se ha aplicado a lenguas mayoritarias y a otros dominios, pero no a la alfabetización bilingüe en lenguas indígenas.

3) La evidencia causal sobre programas en lengua materna en contextos de bajos recursos es prometedora pero escasa y no mide la implementación como una variable continua. En el Perú, un programa de acompañamiento pedagógico a escala en escuelas rurales elevó la lectura en 0,075 desviaciones estándar (Castro et al., 2025), y una intervención de aprendizaje del quechua con apoyo móvil en escuelas EIB produjo efectos de 0,137 a 0,330 desviaciones estándar (Näslund et al., 2026). Fuera de América Latina, un componente de educación bilingüe de transición elevó la alfabetización en la lengua de instrucción en 0,41 y en la lengua materna en 0,75 desviaciones estándar (Leighton, 2022). Estos estudios comparan la presencia o ausencia de un programa; ninguno sigue durante varios años la relación entre la intensidad de la implementación y las trayectorias en ambas lenguas, ni distingue la transferencia intrapersona de las diferencias entre estudiantes y escuelas.

16.1.4 Pregunta, objetivos y contribución

La pregunta que guía el estudio es: ¿cómo se relaciona la intensidad de implementación de la EIB con el nivel inicial y el crecimiento de la

comprensión lectora en lengua originaria y en castellano, y se transfiere, dentro de cada estudiante, el progreso en lengua originaria al castellano? Se persiguen tres objetivos:

a) Establecer que la comprensión lectora se mide con la misma escala en las tres olas y en ambas lenguas;

b) Estimar la asociación de la implementación escolar con el nivel y el crecimiento de ambas trayectorias, y el papel del nivel inicial en lengua originaria como vía de esa asociación; y

c) Estimar la dinámica cruzada intrapersona entre ambas lenguas una vez separadas las diferencias estables.

La contribución es teórica y metodológica, y su alcance excede el caso peruano. En lo teórico, el estudio somete la hipótesis de la interdependencia a una prueba más exigente que la correlacional: la transferencia se evalúa como dinámica intrapersona entre lenguas tipológicamente distantes —el quechua y el aimara, aglutinantes, frente al castellano— en un contexto de alfabetización inicial con recursos escasos. En lo metodológico, combina en un mismo diseño la invarianza longitudinal de la medida, el crecimiento latente en paralelo con un predictor escolar y la separación intrapersona e interpersona, con inferencia que respeta el anidamiento de los estudiantes en escuelas. Las comunidades andinas del Perú son el sitio de prueba; el aporte es un procedimiento y una evidencia transferibles a cualquier sistema que enseñe a leer en una lengua minorizada y en una lengua dominante.

16.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

16.2.1 Interdependencia y transferencia de la comprensión lectora

La comprensión lectora integra la decodificación con el lenguaje oral, el conocimiento previo y procesos de inferencia y control. La hipótesis de la interdependencia propone que estos procesos de orden superior

—identificar la idea principal, inferir relaciones no explícitas, evaluar críticamente lo leído— no se aprenden de nuevo en cada lengua, sino que se apoyan en una competencia subyacente común (Cummins, 1979). La evidencia reciente documenta transferencia de la comprensión lectora y de la composición morfológica entre lenguas tipológicamente distintas, aunque no del vocabulario ni de la lectura de palabras (Xie y Yeung, 2026), y la importancia relativa de las habilidades de cada lengua cambia a lo largo del tiempo (Shen y Huang, 2026). En un sistema que alfabetiza primero en lengua originaria, la transferencia debería operar sobre todo en una dirección: del progreso en lengua originaria al progreso posterior en castellano.

16.2.2 La implementación de la EIB como contexto de aprendizaje

La implementación de la EIB es multidimensional —abarca la organización del programa, la enseñanza cultural y lingüísticamente pertinente, la formación del personal y el vínculo con la comunidad— y en América Latina oscila entre la asimilación y el reconocimiento de los pueblos indígenas (Zsögön, 2025). El dominio docente de la lengua, que el nuevo currículo de formación en EIB del Perú busca fortalecer (Moreano et al., 2026), es una de sus dimensiones más directamente observables. En este estudio la implementación se entiende como la intensidad observable con que la escuela ofrece instrucción en lengua originaria, materiales en esa lengua y docentes que la dominan. Una implementación más intensa debería elevar la lectura en lengua originaria desde el inicio y, por la vía de la transferencia, acelerar después el crecimiento en castellano.

16.2.3 Hipótesis

A partir de ese marco se contrastan cinco hipótesis direccionales:

- **H1.** La intensidad de implementación de la EIB se asocia positivamente con el nivel inicial de comprensión lectora en lengua originaria.
- **H2.** La intensidad de implementación de la EIB se asocia positivamente con el crecimiento anual de la comprensión lectora en castellano.

- **H3.** Dentro de cada estudiante, una comprensión en lengua originaria superior a su propio promedio precede a una comprensión en castellano superior a su propio promedio el año siguiente (efecto cruzado L1→L2 positivo).
- **H4.** El efecto cruzado inverso (castellano→lengua originaria) es menor que el directo.
- **H5.** El nivel inicial en lengua originaria transmite una parte de la asociación entre la implementación y el crecimiento en castellano (efecto indirecto positivo).

16.3 MATERIALES Y MÉTODOS

16.3.1 Diseño

Se realizó un estudio observacional de panel prospectivo con tres olas anuales, al final de los años escolares 2024, 2025 y 2026, sobre una cohorte que cursaba tercer grado de primaria en 2024 y quinto en 2026. El diseño es correlacional: la implementación no se asignó y las asociaciones no se interpretan como efectos causales; la comparación intrapersona controla las diferencias estables entre estudiantes, pero no las variables cambiantes no medidas. Las variables se resumen en la Tabla 16-1.

Tabla 16-1

Operacionalización de los constructos del estudio

Variable	Nivel	Indicadores	Escala	Olas
Comprensión lectora en lengua originaria	Estudiante	Literal; inferencial; crítico-reflexiva	Media de tres subescalas escaladas (cada subescala con media 50 y DE 10 en la ola 1)	2024, 2025, 2026
Comprensión lectora en castellano		Literal; inferencial; crítico-reflexiva	Media de tres subescalas escaladas (cada subescala con media 50 y DE 10 en la ola 1)	

Variable	Nivel	Indicadores	Escala	Olas
Implementación de la EIB	Escuela	Tiempo de instrucción en lengua originaria; materiales en lengua originaria; dominio docente de la lengua	Puntuación factorial estandarizada (media 0, DE 1)	2024
Uso de la lengua originaria en el hogar	Estudiante	Proporción de interacciones familiares en lengua originaria	0 a 1	
Nivel socioeconómico		Índice de bienes, servicios y escolaridad de los padres	Estandarizado	
Sexo		Registro escolar	0 = hombre, 1 = mujer	

Nota: EIB = educación intercultural bilingüe. Las puntuaciones de comprensión de cada ola son la media de las tres subescalas.

16.3.2 Participantes y tamaño de la muestra

La población de referencia fueron las escuelas públicas rurales de primaria registradas como EIB en distritos altoandinos de Apurímac, Ayacucho, Cusco y Puno, con quechua o aimara como lengua originaria predominante. Las escuelas se incorporaron por conveniencia, según su accesibilidad y disposición a participar, de modo que los resultados describen a las escuelas analizadas sin afirmar generalización estadística. Participaron los estudiantes de tercer grado de cada escuela en 2024. El tamaño se fijó con tres criterios, uno por análisis principal, y se adoptó el más exigente (Tabla 16-2). Para el efecto cruzado intrapersona, la potencia se estimó por simulación de Monte Carlo (Muthén y Muthén, 2002), con un efecto focal de 0,10 anclado en la distribución empírica de efectos cruzados (Orth et al., 2024). Para el predictor escolar sobre la pendiente se calculó el número de escuelas con una correlación intraclase de planificación conservadora (Glazerman et al., 2025), y para el ajuste global, con el contraste de ajuste aproximado basado en el RMSEA (MacCallum et al., 1996; Moshagen y

Bader, 2023). Con un desgaste previsto de 20 %, se reclutaron 219 escuelas; la cohorte sumó 2 565 estudiantes (media de 11,7 por escuela; mínimo 6, máximo 24) y 2 165 completaron las tres olas.

Tabla 16-2

Cálculo a priori del tamaño de la muestra

Criterio		Análisis	Parámetros	Ecuación o procedimiento	Resultado
A.	Efecto cruzado intrapersona	Panel con interceptos aleatorios	$\beta = 0{,}10$; estabilidad = 0,40; r intrapersona = 0,30; $\alpha = 0{,}05$; $1 - \beta = 0{,}80$; R = 500	Monte Carlo: potencia = $(1/R)\ \Sigma\ 1[p < \alpha]$	N = 2 100 (potencia 0,846)
B.	Predictor de escuela sobre la pendiente	Crecimiento latente condicional	$\delta = 0{,}20$; $\rho = 0{,}25$; m = 12; $\alpha = 0{,}05$; $1 - \beta = 0{,}80$	$J = (z_1 - \alpha/_2 + z_1 - \beta)^2\ [\rho + (1 - \rho)/m] / \delta^2$	J = 62; N = 744
C.	Ajuste aproximado	Modelo de medida longitudinal	gl = 102; $\varepsilon_0 = 0{,}05$; $\varepsilon_1 = 0{,}08$; $\alpha = 0{,}05$; $1 - \beta = 0{,}80$	χ^2 no central con $\lambda = (N - 1) \cdot gl \cdot \varepsilon^2$	N = 130 (potencia 0,801)
Inflación por desgaste		Todas	Retención prevista = 0,80	$N_1 = [N /$ retención$]$	$N_1 = 2\ 625$
Número de escuelas			m = 12 estudiantes por escuela	$J = [N_1 / m]$	J = 219
Realizado			219 escuelas	Ola 1 / ola 3	2 565 / 2 165 (potencia A interpolada 0,86)

Nota: El tamaño adoptado es el del criterio más exigente (A). N se refiere a estudiantes con las tres olas completas.

16.3.3 Instrumentos y procedimiento

En cada ola cada estudiante resolvió dos pruebas paralelas de comprensión lectora, una en su lengua originaria (quechua o aimara, en la variedad de la escuela) y otra en castellano, equiparadas verticalmente y con tres subescalas (literal, inferencial y crítico-reflexiva) fijadas con media 50 y desviación estándar 10 en la primera ola, de modo que el crecimiento se expresa en una misma métrica. Se aplicaron en las aulas durante las dos

últimas semanas de noviembre, en dos sesiones por lengua con orden contrabalanceado entre escuelas, por aplicadores bilingües formados. La implementación se registró en 2024 mediante observación de dos sesiones de clase, inventario de materiales y registro de dominio docente: porcentaje del tiempo de instrucción en lengua originaria (media 44,9 %, DE 14,0), índice de 0 a 10 de materiales (media 5,33, DE 1,74) y dominio certificado del docente, de 1 a 5 (media 3,04, DE 0,92); un modelo factorial de un factor en las 219 escuelas produjo el índice estandarizado. Las covariables fueron el uso de la lengua originaria en el hogar, el nivel socioeconómico y el sexo.

16.3.4 Estrategia analítica

El análisis siguió cinco pasos:

1) Se describieron la muestra, el desgaste (regresión logística de la permanencia con errores agrupados por escuela) y la correlación intraclase.

2) Un modelo factorial confirmatorio longitudinal de seis factores aportó cargas, fiabilidad, varianza media extraída y la razón heterorrasgo-monorrasgo (HTMT), cuya variante HTMT2 corrige su sesgo (Roemer et al., 2021).

3) Se contrastó la invarianza configural, métrica y escalar en cada lengua con ΔCFI $\leq$ 0,010 y ΔRMSEA $\leq$ 0,015 (Lasker, 2024; Leitgöb et al., 2023).

4) Se compararon modelos de crecimiento sin crecimiento, lineal y de base latente, y se estimó un modelo de crecimiento latente en paralelo con nivel inicial y pendiente de cada lengua regresados en la implementación y las covariables, y la pendiente en castellano regresada además en el nivel inicial en lengua originaria (H1, H2, H5).

5) Se estimó un modelo de panel con interceptos aleatorios (RI-CLPM; Hamaker et al., 2015), con efectos iguales entre intervalos (H3, H4) y en la línea de sus extensiones (Mulder y Hamaker, 2021), se comparó

con el modelo tradicional y se contrastó su robustez con una curva latente con residuos estructurados (Curran et al., 2014).

Los modelos se estimaron por máxima verosimilitud de información completa con los 2 565 estudiantes, apropiada bajo desgaste aleatorio condicionado a lo observado (Tang y Tong, 2025); el ajuste global se evaluó en la submuestra con las tres olas. Los intervalos de los parámetros focales se obtuvieron por bootstrap de percentiles remuestreando escuelas (500 réplicas). Se informan χ^2, CFI, TLI, RMSEA con su intervalo de confianza de 90 % y SRMR, interpretados de forma descriptiva y no contra puntos de corte fijos (Groskurth et al., 2023), considerando la sensibilidad de las pruebas de ajuste en el RI-CLPM (Zheng y Valente, 2023). Los análisis se hicieron con Python 3.12 y semopy 2.3.

16.4 RESULTADOS

16.4.1 Muestra, desgaste y estructura anidada

Completaron la segunda ola 2 360 estudiantes y la tercera 2 165 (84,4 %), con retenciones entre 82,3 % y 86,1 % según la región (Tabla 16-3). La permanencia fue más probable entre quienes tenían mayor comprensión en castellano en la primera ola (OR = 1,23, $p < 0,001$) y mayor nivel socioeconómico (OR = 1,33, $p < 0,001$), y no dependió de la comprensión en lengua originaria ($p = 0,935$), del uso de esa lengua en el hogar ($p = 0,853$) ni del sexo ($p = 0,719$). La correlación intraclase fue de 0,12 a 0,16 en lengua originaria y de 0,19 a 0,24 en castellano, con efectos de diseño entre 2,26 y 3,12 (Tabla 16-4), por debajo de la correlación de planificación de 0,25.

Tabla 16-3

Composición de la muestra y retención por región

Región	Lengua originaria	Escuelas	Ola			Retención	Mujeres	Uso de la lengua originaria en el hogar	NSE
			1	2	3	(%)			
Apurímac	Quechua	37	423	392	364	86,1	52,0	0,67	0,03
Ayacucho	Quechua	38	447	412	368	82,3	53,2	0,66	0,01
Cusco	Quechua	86	1 057	966	893	84,5	49,1	0,67	0,03
Puno	Aimara, Quechua	58	638	590	540	84,6	51,2	0,67	0,00
Total	Quechua, Aimara	219	2 565	2 360	2 165	84,4	50,8	0,67	0,02

Nota: Retención = proporción de estudiantes de la ola 1 que completaron la ola 3. El nivel socioeconómico está estandarizado en la muestra total. Predictores de la permanencia en la ola 3 (regresión logística con errores agrupados por escuela): comprensión en castellano en la ola 1, OR = 1,23, $p < 0,001$; nivel socioeconómico, OR = 1,33, $p < 0,001$; comprensión en lengua originaria, OR = 1,00, $p = 0,935$; uso de la lengua originaria en el hogar, OR = 1,06, $p = 0,853$; sexo, OR = 0,96, $p = 0,719$.

16.4.2 Estadísticos descriptivos, propiedades de la medida e invarianza

La comprensión aumentó de forma sostenida en ambas lenguas, de 49,88 a 53,20 en lengua originaria y de 50,14 a 55,04 en castellano (Tabla 16-4; Figura 16-1, paneles A y B), con distribuciones aproximadamente simétricas (asimetría entre –0,10 y 0,02). La estabilidad entre olas fue alta dentro de cada lengua y la correlación entre lenguas creció de la primera a la tercera ola. El índice de implementación se correlacionó más con las puntuaciones de las olas posteriores que con las de la primera, y el uso de la lengua originaria en el hogar, con la comprensión en esa lengua.

Tabla 16-4

Estadísticos descriptivos, correlación intraclase y correlaciones

	Variable		n	M	DE	ICC	1	2	3	4	5	6	7	8
1.	Lengua originaria	ola 1	2 565	49,88	8,79	0,12	—							
2.		ola 2	2 360	51,58	9,06	0,14	0,78***	—						
3.		ola 3	2 165	53,20	9,33	0,16	0,73***	0,79***	—					
4.	Castellano	ola 1	2 565	50,14	8,79	0,19	0,36***	0,35***	0,35***	—				
5.		ola 2	2 360	52,68	9,39	0,21	0,44***	0,44***	0,44***	0,78***	—			
6.		ola 3	2 165	55,04	9,93	0,24	0,47***	0,49***	0,48***	0,74***	0,80***	—		
7.	Implementación de la EIB		219	0,00	1,00	—	0,07***	0,13***	0,22***	0,10***	0,16***	0,19***	—	
8.	Uso de la lengua originaria en el hogar		2 565	0,67	0,18	0,01	0,23***	0,26***	0,21***	−0,05*	−0,03	0,02	−0,02	—
9.	Nivel socioeconómico		2 565	0,02	0,97	0,00	0,08***	0,07***	0,09***	0,15***	0,14***	0,13***	−0,02	−0,04*

Nota: ICC = correlación intraclase entre escuelas. La implementación de la EIB es un índice de escuela (n = escuelas); su correlación con las variables de los estudiantes se calcula asignando a cada estudiante el valor de su escuela. * p < 0,05; ** p < 0,01; *** p < 0,001.

El modelo de medida longitudinal ajustó bien ($\chi^2(102) = 103{,}11$, $p = 0{,}451$, CFI = 1,000, TLI = 1,000, RMSEA = 0,002 [IC 90 %: 0,000; 0,011], SRMR = 0,008; n = 2 165). Las cargas estandarizadas fueron de 0,73 a 0,89, la fiabilidad ω de 0,85 a 0,88 y la varianza media extraída de 0,65 a 0,72 (Tabla 16-5; Figura 16-1, panel C). La HTMT entre un factor de una lengua y cualquiera de la otra no superó 0,56, y entre olas de una misma lengua osciló entre 0,85 y 0,92, con los menores valores entre la primera y la tercera ola. El índice escolar de implementación tuvo cargas de 0,56 a 0,71 y una fiabilidad ω = 0,67; como el modelo de tres indicadores está exactamente identificado, su ajuste no es contrastable.

Tabla 16-5

Modelo de medida longitudinal y validez discriminante

Factor		λ **literal**	λ inferencial	λ crítico-reflexiva	ω	α	AVE	HTMT máxima con la otra lengua
Comprensión lectora en lengua originaria	ola 1	0,82	0,87	0,73	0,85	0,85	0,65	0,55
	ola 2	0,81	0,87	0,75	0,85	0,85	0,66	0,56
	ola 3	0,84	0,87	0,77	0,87	0,87	0,68	0,55
Comprensión lectora en castellano	ola 1	0,83	0,86	0,74	0,85	0,85	0,66	0,43
	ola 2	0,84	0,87	0,76	0,87	0,86	0,68	0,52
	ola 3	0,86	0,89	0,79	0,88	0,88	0,72	0,56
Implementación de la EIB (escuelas, J = 219)		0,71[a]	0,64[b]	0,56[c]	0,67	—	0,41	—

Nota: λ = carga estandarizada; AVE = varianza media extraída; HTMT = razón heterorrasgo-monorrasgo. Ajuste del modelo de seis factores: $\chi^2(102) = 103{,}11$, $p = 0{,}451$, CFI = 1,000, TLI = 1,000, RMSEA = 0,002 [IC 90 %: 0,000; 0,011], SRMR = 0,008. En la fila de la implementación, las tres columnas de carga corresponden a [a] tiempo de instrucción en lengua originaria, [b] materiales en lengua originaria y [c] dominio docente de la lengua.

Tabla 16-6

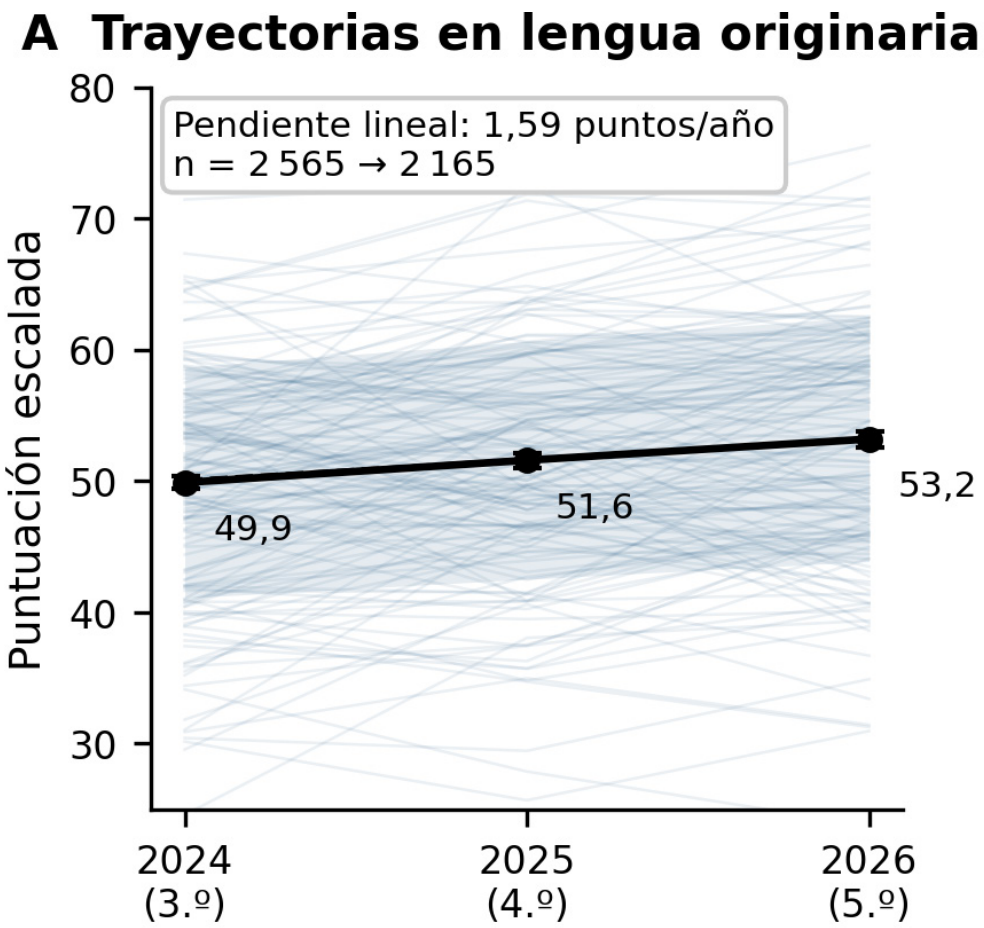
Invarianza longitudinal de la medida de comprensión lectora

Lengua	Modelo	χ^2	gl	p	CFI	TLI	RMSEA [IC 90 %]	SRMR	$\Delta\chi^2$ (Δgl)	ΔCFI	ΔRMSEA
Originaria	Configural	17,03	15	0,317	1,000	1,000	0,008 [0,000; 0,022]	0,005	—	—	—
	Métrica	22,25	19	0,272	1,000	1,000	0,009 [0,000; 0,022]	0,013	5,22 (4)	0,000	0,001
	Escalar	27,25	23	0,246	1,000	1,000	0,009 [0,000; 0,021]	0,013	4,99 (4)	0,000	0,000
Castellano	Configural	6,85	15	0,962	1,000	1,001	0,000 [0,000; 0,000]	0,004	—	—	—
	Métrica	7,42	19	0,992	1,000	1,002	0,000 [0,000; 0,000]	0,005	0,57 (4)	0,000	0,000
	Escalar	22,13	23	0,512	1,000	1,000	0,000 [0,000; 0,017]	0,006	14,71 (4)	0,000	0,000

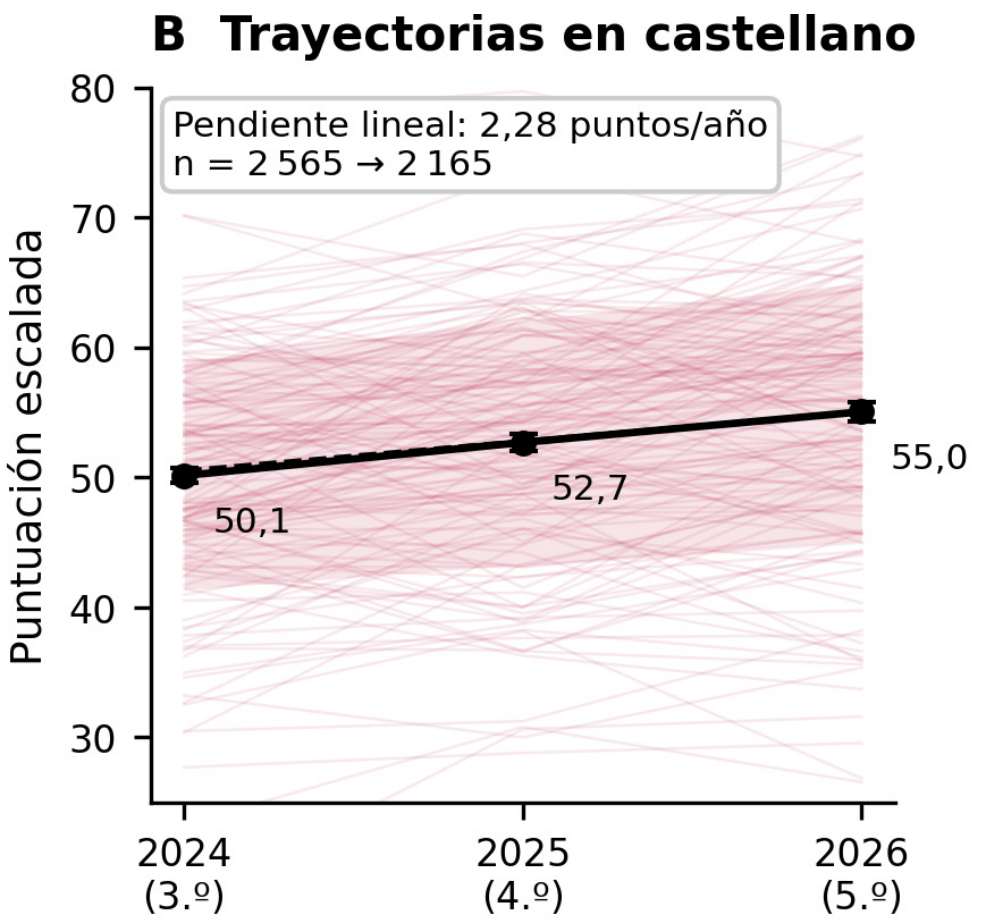
Nota: n = 2 165 estudiantes con las tres olas. Cada modelo incluye las covarianzas residuales de la misma subescala entre olas. Los incrementos se calculan respecto del modelo inmediatamente menos restringido.

Figura 16-1

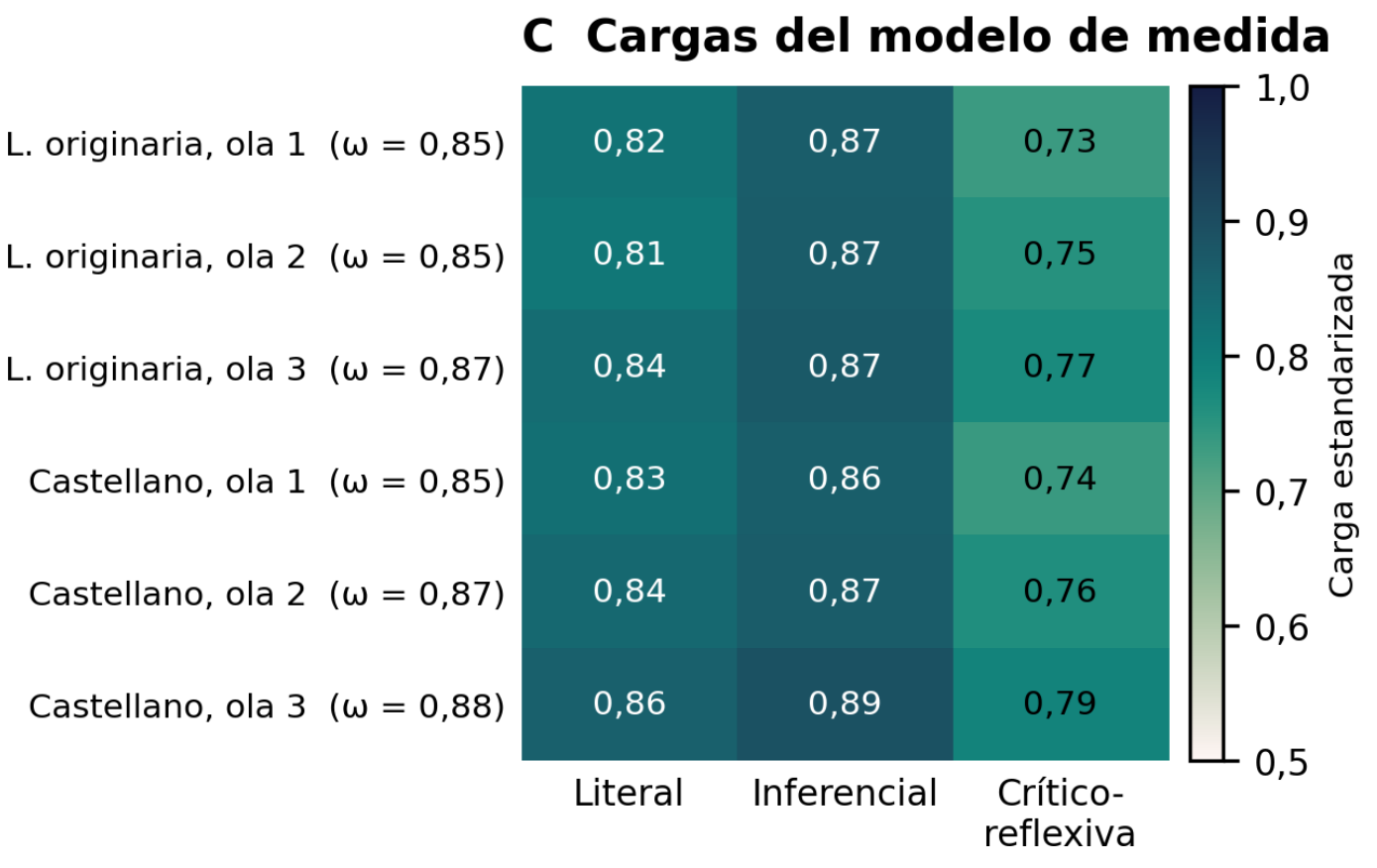
Trayectorias observadas de la comprensión lectora y propiedades de la medida



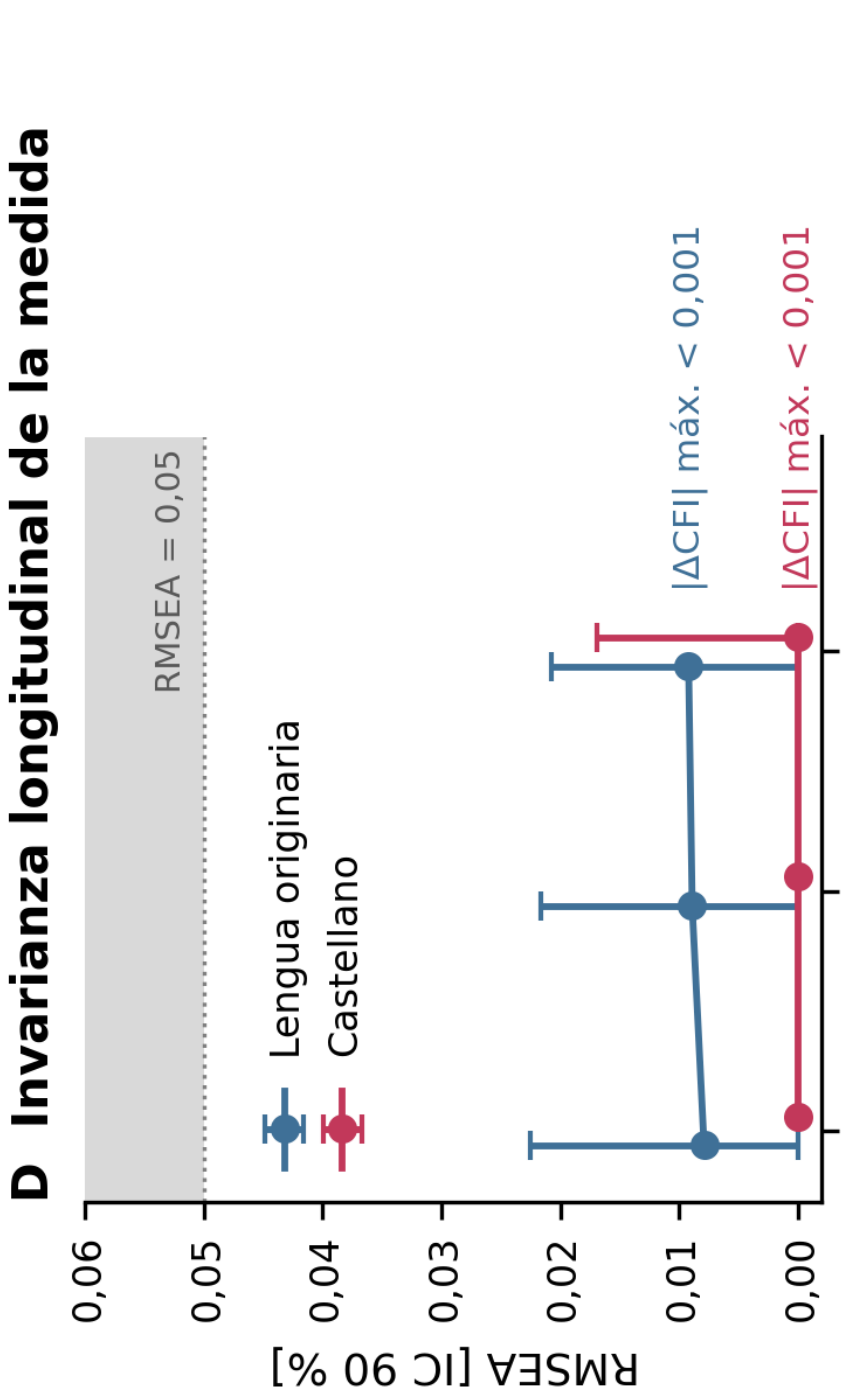
Nota. Panel de tres olas (2024-2026) con 2 565 estudiantes de 219 escuelas en la ola 1 y 2 165 en la ola 3. Paneles A y B: trayectorias individuales de 150 estudiantes con las tres olas elegidos al azar (líneas finas), banda de ±1 DE, media observada por ola con IC 95 % corregido por el efecto de diseño (puntos negros) y trayectoria del modelo de crecimiento lineal (línea discontinua); la pendiente media fue de 1,59 puntos por año en lengua originaria y de 2,28 en castellano.



Nota. Panel de tres olas (2024-2026) con 2 565 estudiantes de 219 escuelas en la ola 1 y 2 165 en la ola 3. Paneles A y B: trayectorias individuales de 150 estudiantes con las tres olas elegidos al azar (líneas finas), banda de ±1 DE, media observada por ola con IC 95 % corregido por el efecto de diseño (puntos negros) y trayectoria del modelo de crecimiento lineal (línea discontinua); la pendiente media fue de 1,59 puntos por año en lengua originaria y de 2,28 en castellano.



Nota. Panel de tres olas (2024-2026) con 2 565 estudiantes de 219 escuelas en la ola 1 y 2 165 en la ola 3. Panel C: cargas estandarizadas del modelo factorial longitudinal de seis factores (n = 2 165) y fiabilidad ω de cada factor; ajuste $\chi^2(102) = 103{,}11$, CFI = 1,000, RMSEA = 0,002, SRMR = 0,008.



Nota. Panel de tres olas (2024-2026) con 2 565 estudiantes de 219 escuelas en la ola 1 y 2 165 en la ola 3. Panel D: RMSEA con IC 90 % de los modelos configural, métrico y escalar en cada lengua; el mayor cambio de CFI entre niveles fue < 0,001 en lengua originaria y < 0,001 en castellano

La invarianza escalar se sostuvo en ambas lenguas: imponer cargas e interceptos iguales entre olas apenas modificó el ajuste (Tabla 16-6; Figura 16-1, panel D). Con la escala común, las medias latentes aumentaron en lengua originaria 1,69 y 3,19 puntos y en castellano 2,44 y 4,70 puntos respecto de la primera ola.

16.4.3 Trayectorias de crecimiento

En las dos lenguas, el modelo sin crecimiento fue inadmisible y el lineal ajustó sin desajuste detectable (Tabla 16-7). El modelo de base latente estimó la carga de la tercera ola en 1,93 en ambas lenguas, próxima al valor 2 del crecimiento lineal, y no mejoró el ajuste de forma significativa; se retuvo el modelo lineal. La pendiente media fue de 1,59 puntos por año en lengua originaria (0,18 DE de la primera ola) y de 2,28 en castellano (0,26 DE).

El modelo de base latente estimó la carga de la tercera ola en 1,93 en ambas lenguas, próxima al valor 2 del crecimiento lineal, y no mejoró el ajuste de forma significativa; se retuvo el modelo lineal. La pendiente media fue de 1,59 puntos por año en lengua originaria (0,18 DE de la primera ola) y de 2,28 en castellano (0,26 DE).

16.4.4 Implementación de la EIB, nivel inicial y crecimiento (H1, H2 y H5)

El modelo de crecimiento latente en paralelo condicional ajustó bien ($\chi^2(15) = 23{,}16$, $p = 0{,}081$, CFI = 0,999, TLI = 0,999, RMSEA = 0,016 [IC 90 %: 0,000; 0,028], SRMR = 0,009; Tabla 16-8). Una desviación estándar más de implementación se asoció con un nivel inicial mayor en lengua originaria ($b = 0{,}64$, $\beta = 0{,}08$, IC 95 % por bootstrap de escuelas [0,12; 1,17]; equivalente a 0,07 DE), lo que respalda H1. La misma diferencia se asoció con un crecimiento mayor en castellano, de 0,44 puntos por año ($\beta = 0{,}23$, IC 95 % [0,27; 0,60]; 0,050 DE por año y 0,10 DE en los dos años), lo que respalda H2, y con un crecimiento mayor en lengua originaria ($b = 0{,}69$, $\beta = 0{,}35$, IC 95 % [0,56; 0,84]). El nivel inicial en lengua originaria predijo el crecimiento en castellano ($\beta = 0{,}42$), y el efecto indirecto de la

Tabla 16-7

ÍComparación de modelos de crecimiento latente por lengua

Lengua	Modelo	χ^2	gl	p	CFI	RMSEA [IC 90 %]	SRMR	Nivel inicial medio	Pendiente media	Diferencia con el lineal: $\Delta\chi^2$ (Δgl); p
Oiginaria	Sin crecimiento	505,89	4	< 0,001	0,884	0,241 [0,223; 0,259]	0,073	51,61	—	505,64 (3); p < 0,001
	Lineal	0,26	1	0,612	1,000	0,000 [0,000; 0,045]	0,000	50,04	1,59	—
	Base latente	0,00	0	—	1,000	0,000 [0,000; 0,000]	0,000	50,03	1,64	0,26 (1); p = 0,612
Castellano	Sin crecimiento	944,98	4	< 0,001	0,790	0,330 [0,312; 0,348]	0,119	52,75	—	944,31 (3); p < 0,001
	Lineal	0,66	1	0,415	1,000	0,000 [0,000; 0,053]	0,000	50,51	2,28	—
	Base latente	0,00	0	—	1,000	0,000 [0,000; 0,000]	0,000	50,48	2,37	0,66 (1); p = 0,415

Nota: n = 2 165. En el modelo de base latente la carga de la tercera ola se estima libremente (1,93 en lengua originaria y 1,93 en castellano), por lo que el modelo queda saturado (gl = 0). Las medias están en la métrica de la puntuación escalada.

implementación por esa vía fue ab = 0,065 (IC 95 % [0,013; 0,117]), lo que respalda H5.

Tabla 16-8

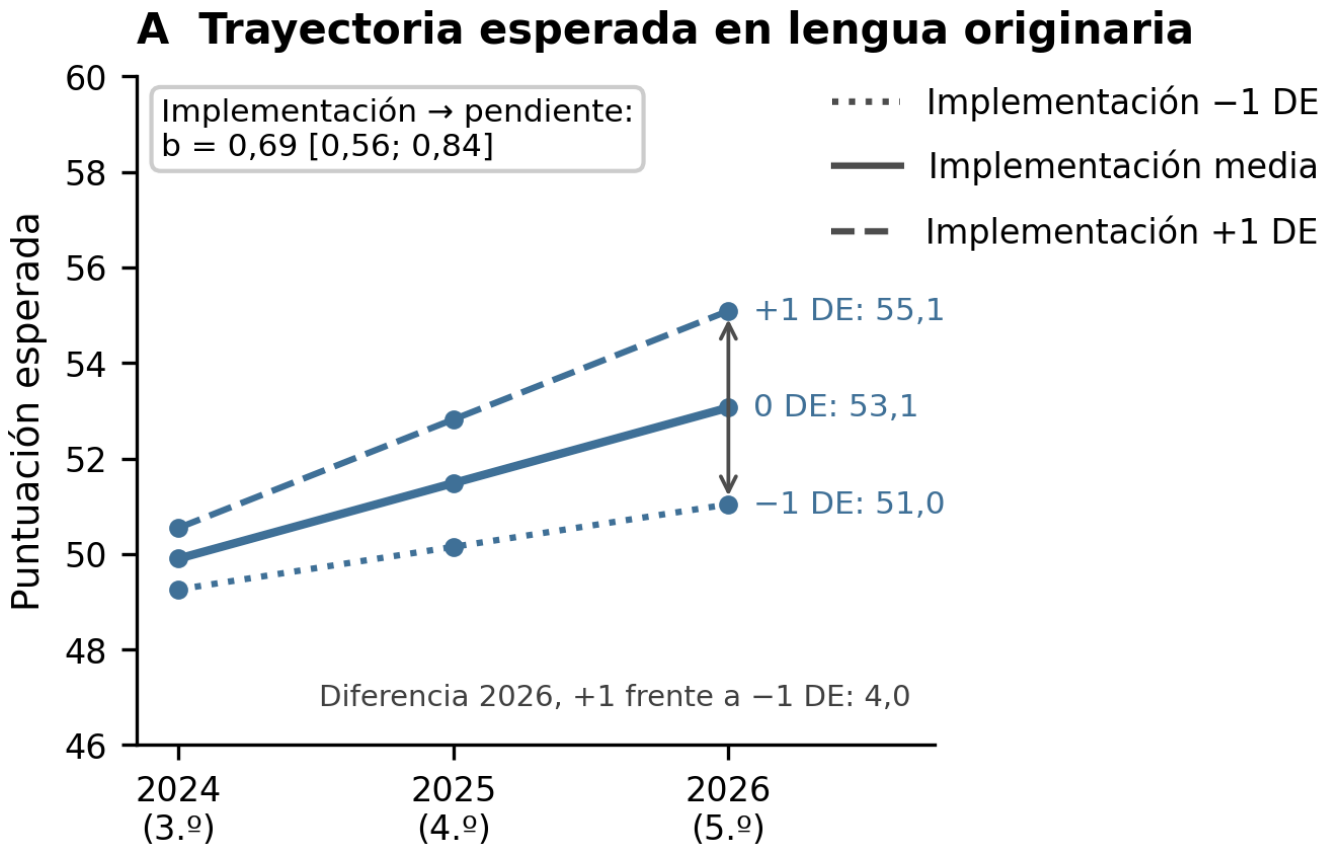
Modelo de medida longitudinal y validez discriminante

Resultado	Predictor	b	EE	β	p	IC 95 % bootstrap
Nivel inicial, lengua originaria	Implementación de la EIB	0,643	0,166	0,08	< 0,001	[0,116; 1,167]
	Uso de la lengua originaria en el hogar	2,126	0,166	0,27	< 0,001	—
	Nivel socioeconómico	0,799	0,166	0,10	< 0,001	—
	Mujer	0,074	0,331	0,00	0,823	—
Pendiente, lengua originaria	Implementación de la EIB	0,691	0,065	0,35	< 0,001	[0,561; 0,840]
	Uso de la lengua originaria en el hogar	0,024	0,065	0,01	0,716	—
	Nivel socioeconómico	0,063	0,065	0,03	0,331	—
	Mujer	0,122	0,130	0,03	0,344	—
Nivel inicial, castellano	Implementación de la EIB	0,933	0,169	0,12	< 0,001	—
	Uso de la lengua originaria en el hogar	–0,385	0,169	–0,05	0,022	—
	Nivel socioeconómico	1,359	0,169	0,17	< 0,001	—
	Mujer	0,122	0,337	0,01	0,716	—
Pendiente, castellano	Implementación de la EIB	0,438	0,065	0,23	< 0,001	[0,272; 0,600]
	Uso de la lengua originaria en el hogar	0,090	0,068	0,05	0,184	—
	Nivel socioeconómico	–0,057	0,065	–0,03	0,386	—
	Mujer	0,108	0,130	0,03	0,406	—
	Nivel inicial, lengua originaria	0,101	0,010	0,42	< 0,001	[0,078; 0,126]
	Efecto indirecto de la implementación por el nivel inicial en lengua originaria	0,065	0,026	—	0,016	[0,013; 0,117]

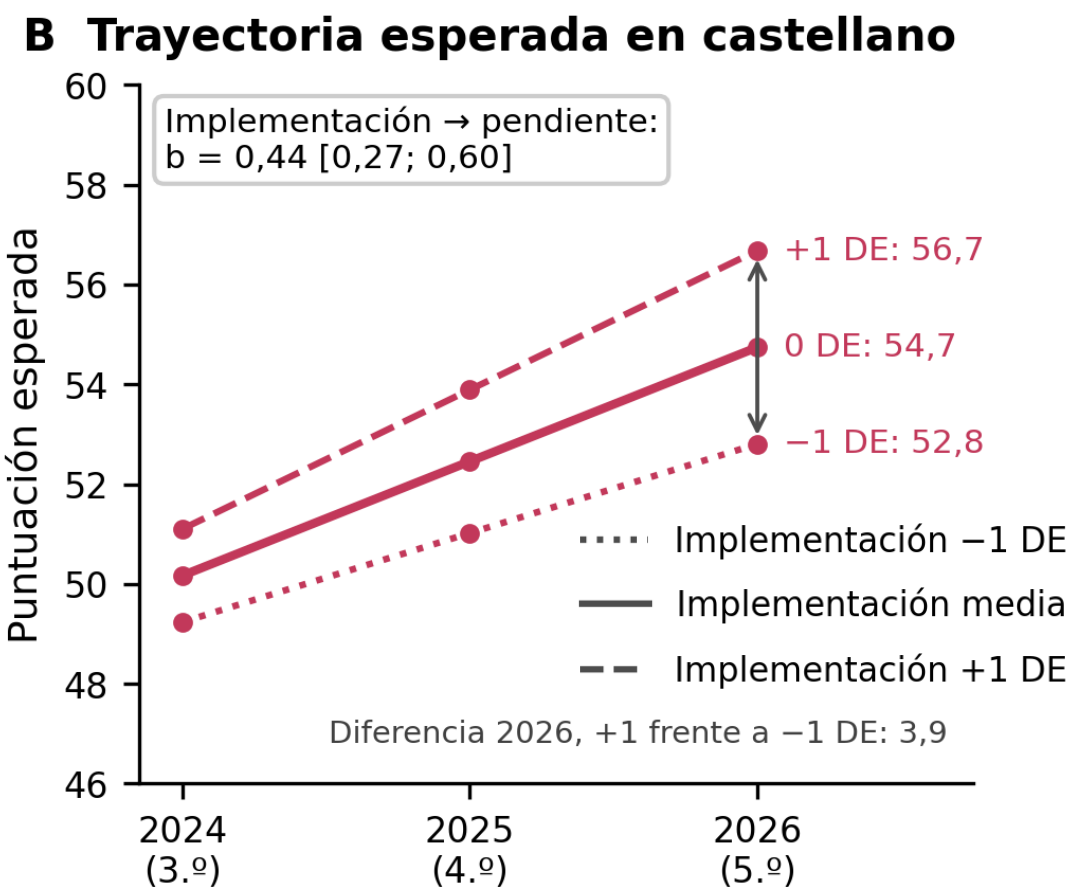
Nota: N = 2 565 (información completa). b = coeficiente no estandarizado en puntos de la escala (en las pendientes, puntos por año); β = coeficiente estandarizado; EE = error estándar del modelo. Los intervalos de confianza se obtuvieron por bootstrap de percentiles remuestreando escuelas (500 réplicas); el error estándar y el valor p del efecto indirecto proceden del bootstrap. El uso de la lengua originaria en el hogar y el nivel socioeconómico están estandarizados. Ajuste: $\chi^2(15) = 23{,}16$, p = 0,081, CFI = 0,999, TLI = 0,999, RMSEA = 0,016 [IC 90 %: 0,000; 0,028], SRMR = 0,009.

Figura 16-2

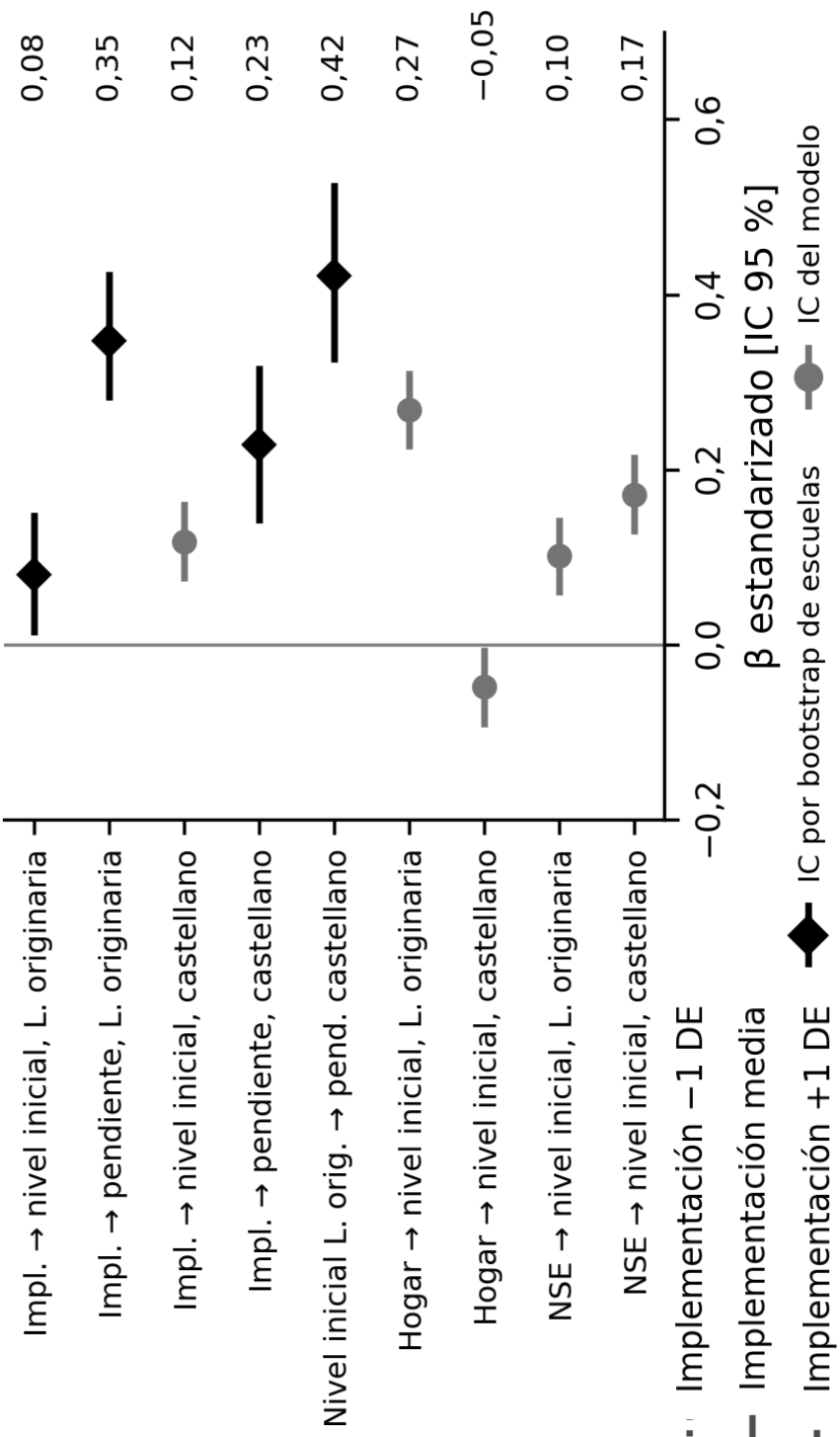
Implementación de la educación intercultural bilingüe y trayectorias de comprensión lectora



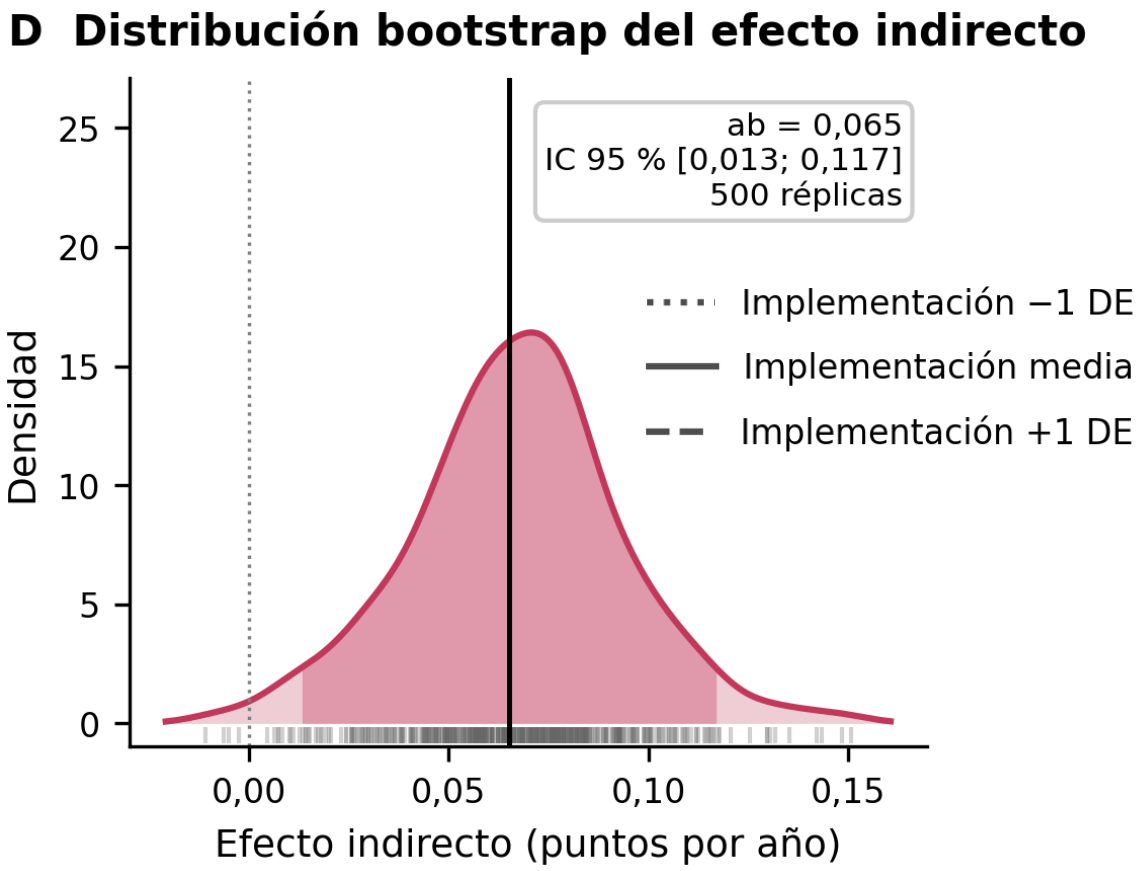
Nota. Modelo de crecimiento latente en paralelo condicional estimado con información completa (N = 2 565); intervalos de confianza por bootstrap de percentiles remuestreando escuelas (500 réplicas). Paneles A y B: trayectorias esperadas para escuelas con implementación 1 DE por debajo de la media, en la media y 1 DE por encima, con las covariables en su media; una DE más de implementación se asoció con una pendiente mayor en lengua originaria (b = 0,69, IC 95 % [0,56; 0,84]) y en castellano (b = 0,44, IC 95 % [0,27; 0,60]).



Nota. Modelo de crecimiento latente en paralelo condicional estimado con información completa (N = 2 565); intervalos de confianza por bootstrap de percentiles remuestreando escuelas (500 réplicas). Paneles A y B: trayectorias esperadas para escuelas con implementación 1 DE por debajo de la media, en la media y 1 DE por encima, con las covariables en su media; una DE más de implementación se asoció con una pendiente mayor en lengua originaria (b = 0,69, IC 95 % [0,56; 0,84]) y en castellano (b = 0,44, IC 95 % [0,27; 0,60]).



Nota. Panel C: coeficientes estandarizados con IC 95 %; los de la implementación sobre el nivel inicial en lengua originaria (β = 0,08) y sobre las pendientes, y el del nivel inicial en lengua originaria sobre la pendiente en castellano (β = 0,42), con IC por bootstrap; el resto, con IC del modelo.



Nota. Panel D: distribución bootstrap del efecto indirecto de la implementación sobre la pendiente en castellano por la vía del nivel inicial en lengua originaria (ab = 0,065, IC 95 % [0,013; 0,117]); la zona oscura es el IC 95 %.

El uso de la lengua originaria en el hogar se asoció con un nivel inicial mayor en esa lengua ($\beta = 0,27$) y menor en castellano ($\beta = -0,05$, $p = 0,022$), pero no con el crecimiento de ninguna de las dos; el nivel socioeconómico se asoció con los niveles iniciales y no con las pendientes, y el sexo no se asoció con ningún componente.

16.4.5 Dinámica intrapersona entre lenguas (H3 y H4)

El modelo de panel con interceptos aleatorios ajustó de forma aceptable ($\chi^2(5) = 52,36$, $p < 0,001$, CFI = 0,995, TLI = 0,985, RMSEA = 0,066 [IC 90 %: 0,051; 0,083], SRMR = 0,030) y mejor que el modelo de panel tradicional ($\chi^2(8) = 392,16$, $p < 0,001$, CFI = 0,960, TLI = 0,925, RMSEA = 0,149 [IC 90 %: 0,137; 0,162], SRMR = 0,034; $\Delta\chi^2(3) = 339,79$, $p < 0,001$). Las diferencias estables entre estudiantes explicaron el 63,6 % de la varianza de la comprensión en lengua originaria y el 71,9 % en castellano, y los interceptos aleatorios de ambas lenguas se correlacionaron r = 0,54 (Tabla 16-9; Figura 16-3). Dentro de cada estudiante, una comprensión en lengua originaria superior a su propio promedio precedió a una comprensión en castellano superior a su propio promedio el año siguiente ($b = 0,266$, $\beta = 0,24$ y 0,26 en los dos intervalos), lo que respalda H3. El efecto inverso también fue positivo, pero menor ($b = 0,159$, $\beta = 0,11$ y 0,14); la diferencia entre ambos fue 0,107 (IC 95 % [0,056; 0,153]), lo que respalda H4. El efecto de la lengua originaria sobre el castellano duplicó el percentil 75 de la distribución empírica de efectos cruzados en modelos con interceptos aleatorios (0,11; Orth et al., 2024), y el inverso se situó en torno a él. El modelo tradicional estimó una estabilidad de $b = 0,77$ en ambas lenguas y efectos cruzados distintos ($b = 0,185$ y 0,097).

La curva latente con residuos estructurados ajustó prácticamente sin desajuste ($\chi^2(1) = 1,00$, $p = 0,317$, CFI = 1,000, TLI = 1,000, RMSEA = 0,001 [IC 90 %: 0,000; 0,057], SRMR = 0,003) y reprodujo la dirección y el orden de los efectos cruzados (L1→L2: $b = 0,206$, $p = 0,057$; L2→L1: $b = 0,161$, $p = 0,121$), con menor precisión. La varianza de la pendiente individual en lengua originaria fue prácticamente nula en este modelo (< 0,001).

Tabla 16-9

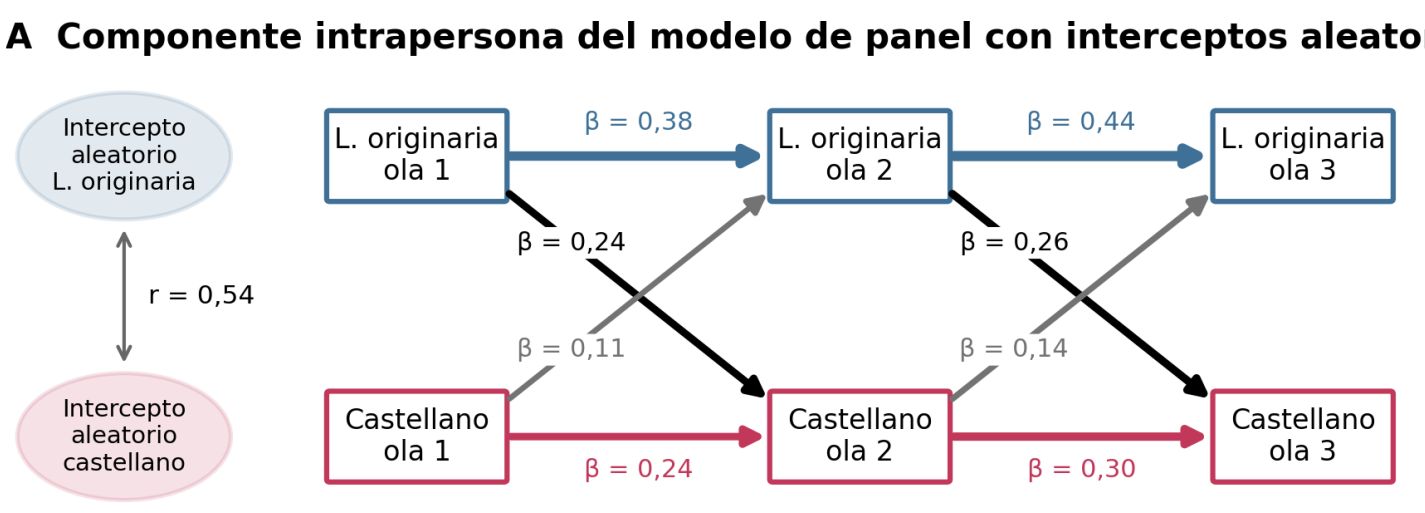
Dinámica intrapersona entre la comprensión en lengua originaria y en castellano

Efecto	RI-CLPM: b	EE	β (ola 1→2; 2→3)	p	IC 95 % bootstrap	CLPM: b	LCM-SR:	
							b	p
Autorregresivo, lengua originaria	0,456	0,030	0,38; 0,44	< 0,001	[0,383; 0,528]	0,770	0,270	0,055
Autorregresivo, castellano	0,325	0,033	0,24; 0,30	< 0,001	[0,259; 0,385]	0,768	0,116	0,417
Cruzado lengua originaria → castellano (H3)	0,266	0,026	0,24; 0,26	< 0,001	[0,209; 0,317]	0,185	0,206	0,057
Cruzado castellano → lengua originaria (H4)	0,159	0,027	0,11; 0,14	< 0,001	[0,102; 0,215]	0,097	0,161	0,121
Diferencia de efectos cruzados (H3 – H4)	0,107	—	—	—	[0,056; 0,153]	0,088	0,044	—
Ajuste: χ^2 (gl)	52,36 (5)	—	—	—	—	392,16 (8)	1,00 (1)	—
Ajuste: CFI; RMSEA	0,995; 0,066	—	—	—	—	0,960; 0,149	1,000; 0,001	—

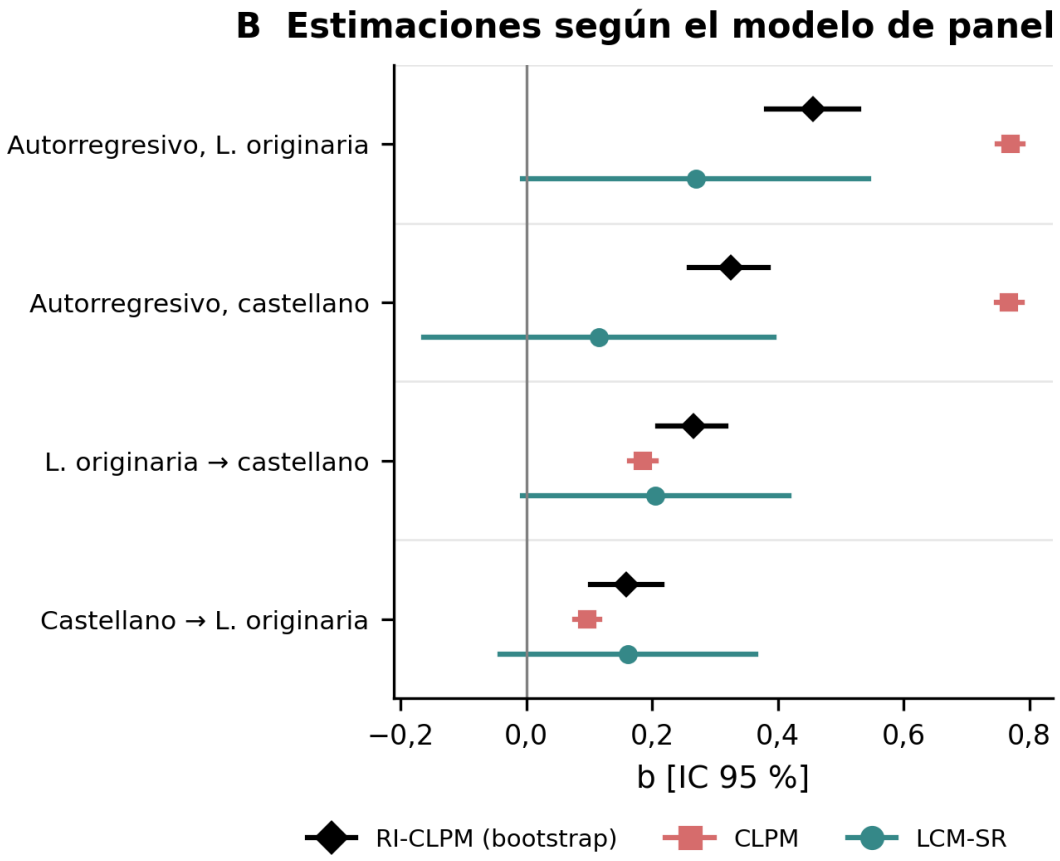
Nota: RI-CLPM = modelo de panel con interceptos aleatorios, estimado con información completa (N = 2 565); CLPM = modelo de panel tradicional; LCM-SR = curva latente con residuos estructurados; los dos últimos y los índices de ajuste se estimaron con los 2 165 estudiantes con las tres olas. Los efectos se restringieron a ser iguales entre intervalos; el β difiere entre intervalos porque cambian las varianzas. Intervalos de confianza por bootstrap de escuelas (500 réplicas).

Figura 16-3

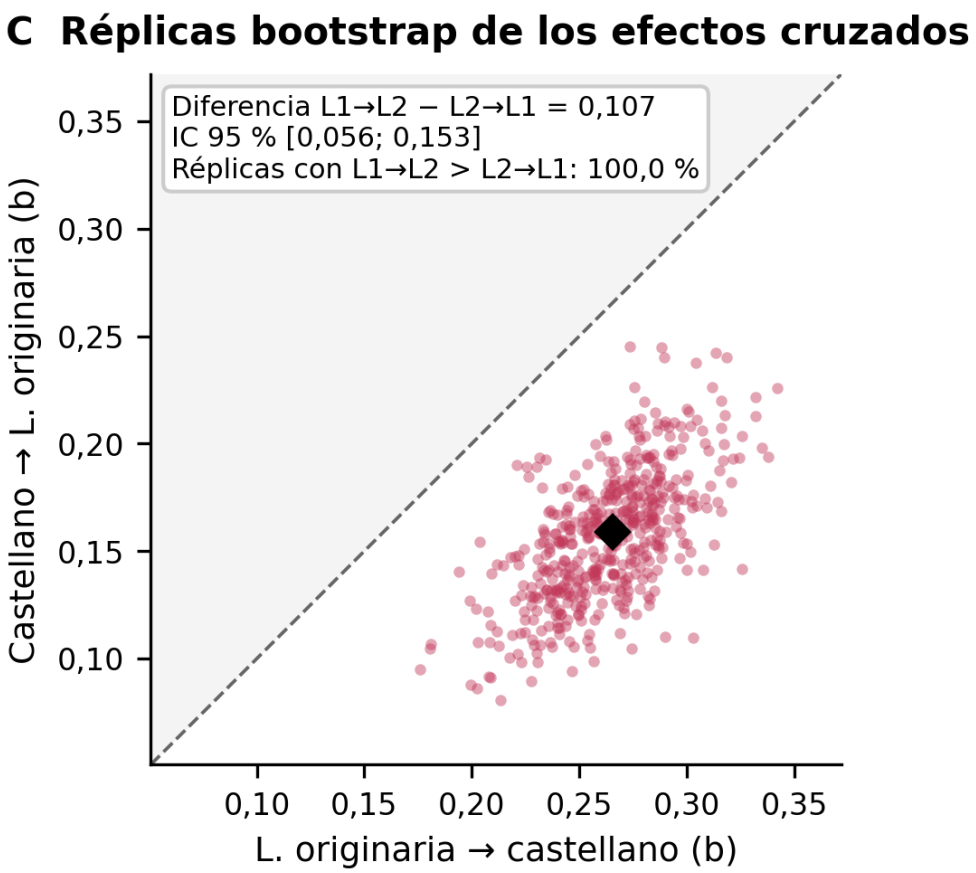
Dinámica intrapersona entre la comprensión lectora en lengua originaria y en castellano



Nota. Panel A: componente intrapersona del modelo de panel con interceptos aleatorios (RI-CLPM) estimado con información completa (N = 2 565), con efectos autorregresivos y cruzados iguales entre intervalos; se muestran los coeficientes estandarizados de cada intervalo y la correlación entre interceptos aleatorios (r = 0,54); ajuste $\chi^2(5) = 52,36$, CFI = 0,995, RMSEA = 0,066.



Nota. Panel B: coeficientes no estandarizados con IC 95 % en el RI-CLPM (IC por bootstrap de escuelas, 500 réplicas), en el modelo de panel tradicional (CLPM; $\chi^2(8) = 392{,}16$) y en la curva latente con residuos estructurados (LCM-SR; $\chi^2(1) = 1{,}00$), estos dos con IC del modelo sobre los 2 165 estudiantes con las tres olas.



Nota. Panel C: réplicas bootstrap de los dos efectos cruzados del RI-CLPM; el rombo es la estimación de la muestra completa (L1→L2: b = 0,266; L2→L1: b = 0,159) y la zona sombreada, la región en que el efecto inverso superaría al directo; diferencia 0,107, IC 95 % [0,056; 0,153]. L1 = lengua originaria; L2 = castellano.

16.4.6 Análisis de robustez

Los resultados centrales se mantuvieron con la submuestra de casos completos y al separar las escuelas quechuas de las aimaras (Tabla 10). La asociación de la implementación con el crecimiento en castellano fue positiva en todas las especificaciones (b de 0,35 a 0,48), igual que la del nivel inicial en lengua originaria con ese crecimiento, y el efecto cruzado L1→L2 superó al inverso en todas ellas.

Tabla 16-10

Análisis de robustez de los efectos focales

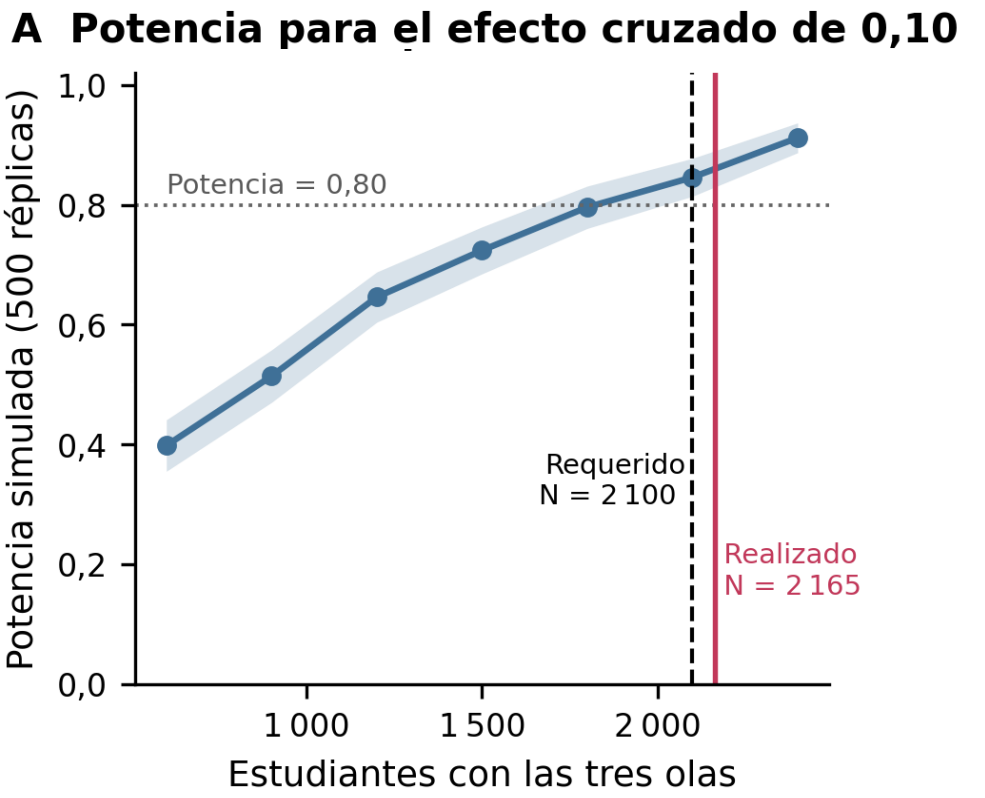
Especificación	n	Implementación		Nivel inicial en lengua originaria → pendiente en castellano:	Cruzado	
		→ pendiente en castellano:	→ nivel inicial en lengua originaria:		L1→L2:	L2→L1:
		b (p)				
Modelos principales (información completa)	2 565	0,438 (< 0,001)	0,643 (< 0,001)	0,101 (< 0,001)	0,266 (< 0,001)	0,159 (< 0,001)
Solo casos completos	2 165	0,478 (< 0,001)	0,669 (< 0,001)	0,098 (< 0,001)	0,266 (< 0,001)	0,164 (< 0,001)
Solo escuelas quechuas	2 164	0,447 (< 0,001)	0,663 (< 0,001)	0,103 (< 0,001)	0,282 (< 0,001)	0,161 (< 0,001)
Solo escuelas aimaras	401	0,354 (0,048)	0,579 (0,207)	0,096 (< 0,001)	0,167 (0,011)	0,127 (0,052)
Error estándar: modelo / bootstrap de escuelas	—	0,065 / 0,084	0,166 / 0,258	0,010 / 0,012	0,026 / 0,027	0,027 / 0,030

Nota: Las especificaciones de casos completos se estimaron por máxima verosimilitud sobre los estudiantes con las tres olas; las de subgrupos, con información completa dentro de cada subgrupo. La lengua corresponde a la lengua originaria de la escuela.

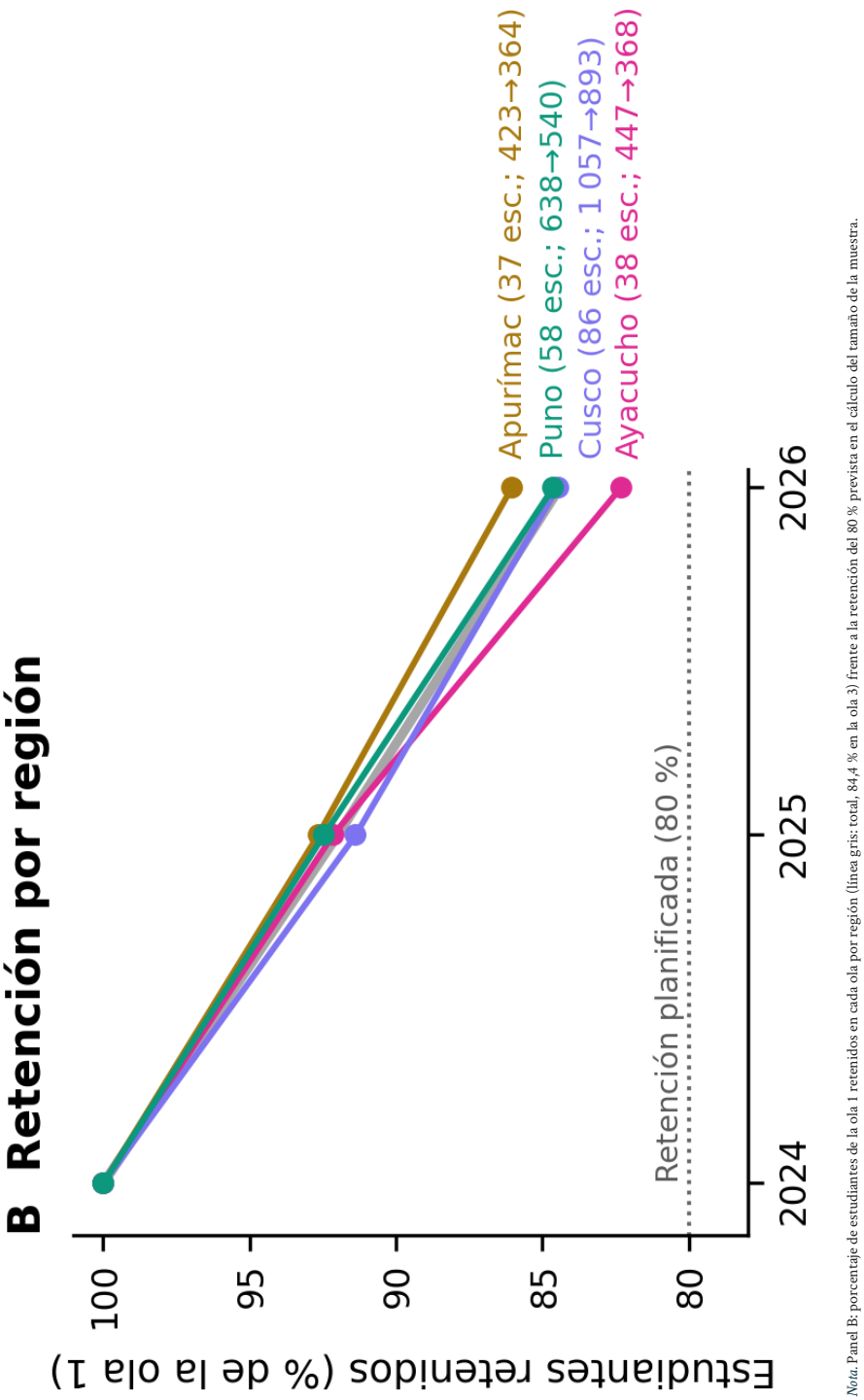
leta dentro de cada subgrupo. La lengua corresponde a la lengua originaria de la escuela.

Figura 16-4

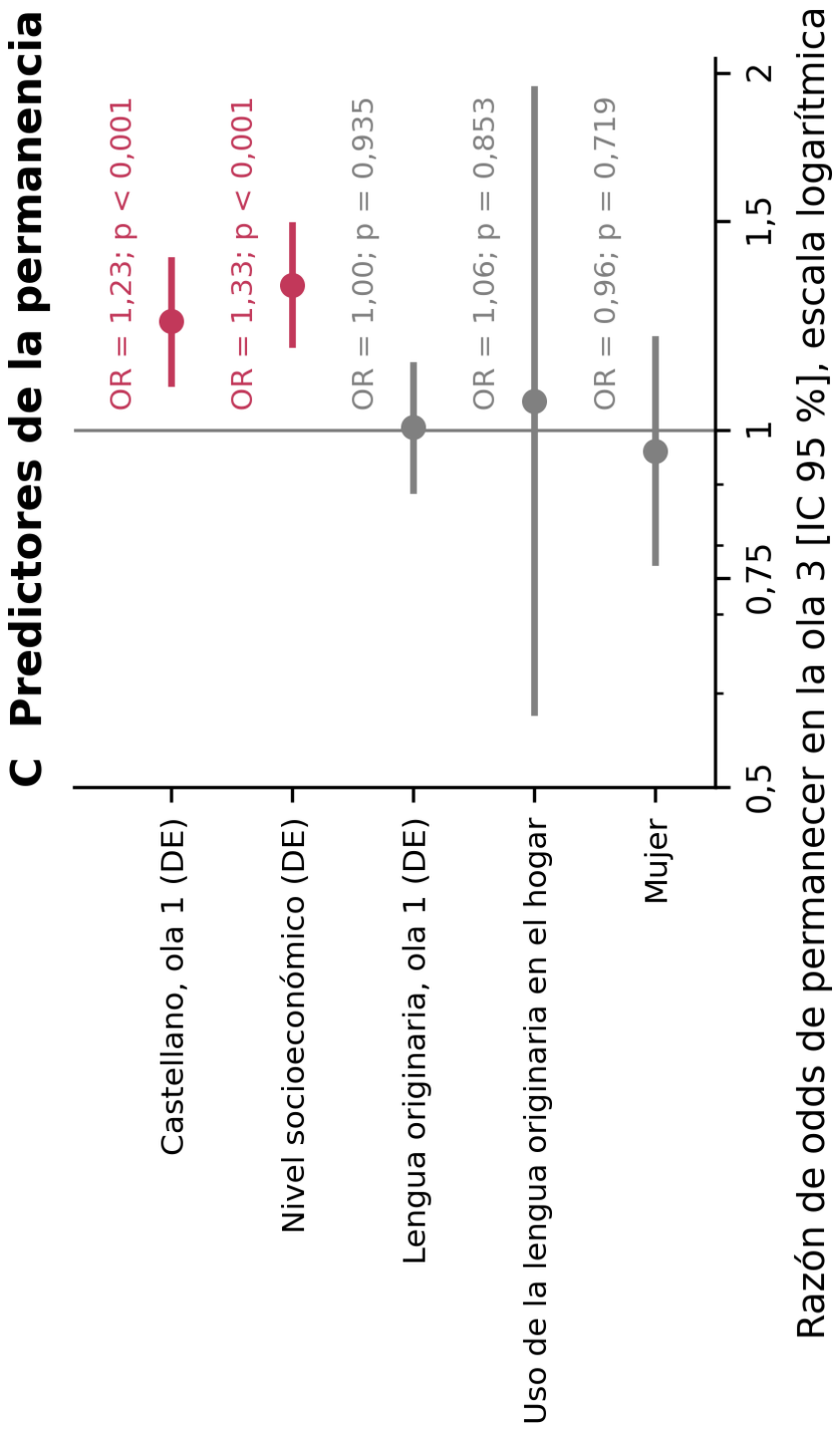
Potencia del diseño, retención y estructura anidada de la muestra



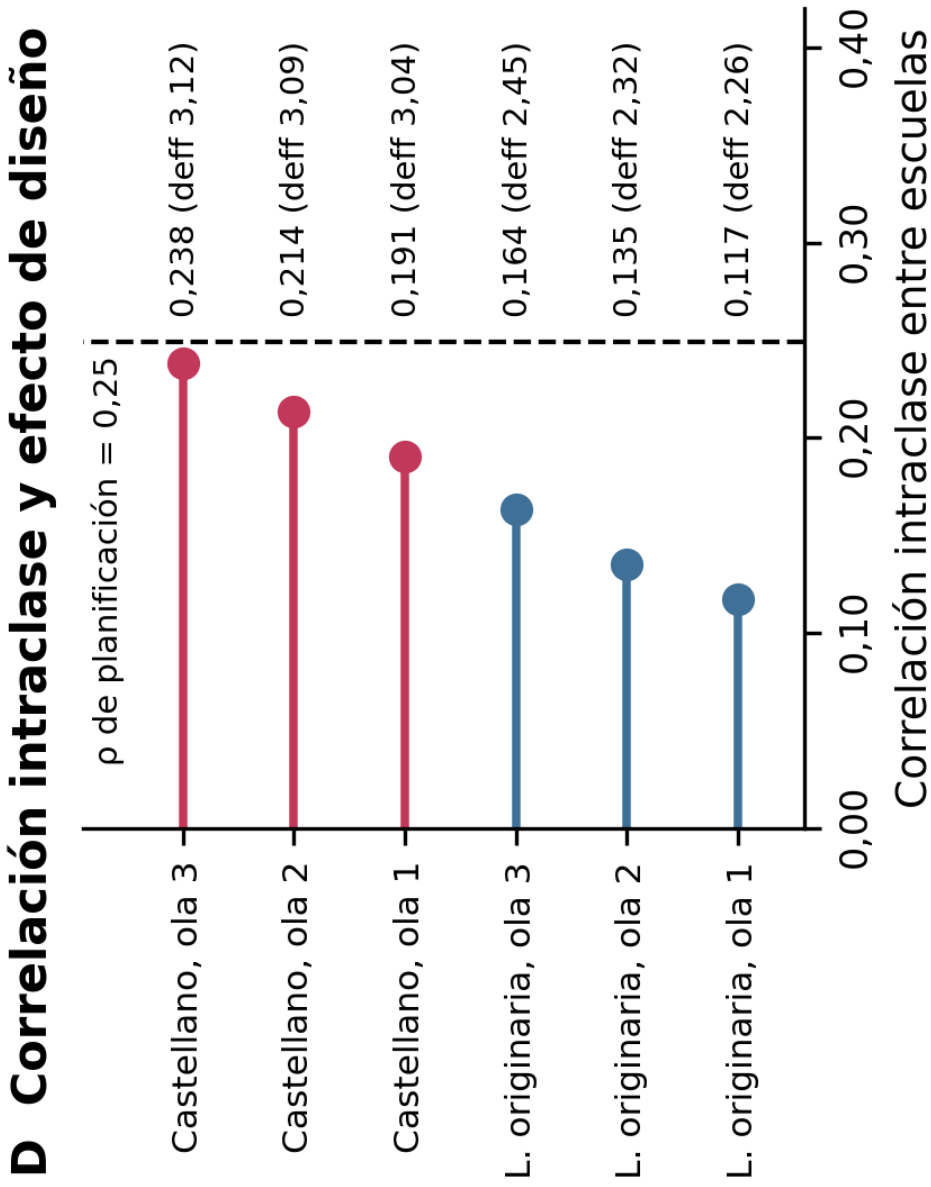
Nota. Panel A: potencia estimada por Monte Carlo (500 réplicas por tamaño, banda = IC 95 % binomial) para detectar un efecto cruzado intrapersona estandarizado de 0,10 con $\alpha = 0,05$ en un modelo de panel con interceptos aleatorios de tres olas; el cálculo a priori exigió 2 100 estudiantes con las tres olas y se completaron 2 165 (potencia interpolada 0,86).



Nota. Panel B: porcentaje de estudiantes de la ola 1 retenidos en cada ola por región (línea gris: total, 84,4 % en la ola 3) frente a la retención del 80 % prevista en el cálculo del tamaño de la muestra.



Nota. Panel C: razones de odds de completar la ola 3 con IC 95 % (regresión logística con errores agrupados por escuela); la permanencia fue más probable con mayor comprensión en castellano en la ola 1 (OR = 1,23, p < 0,001) y mayor nivel socioeconómico (OR = 1,33, p < 0,001).



Nota. Panel D: correlación intraclase entre escuelas y efecto de diseño (deff) de cada puntuación frente a la correlación de planificación ρ = 0,25 usada en el cálculo del número de escuelas..

En las 36 escuelas aimaras, con menos estudiantes, los efectos conservaron su signo con intervalos más amplios; la asociación de la implementación con el nivel inicial en lengua originaria no alcanzó significación en ese subgrupo (p = 0,207). Los errores estándar del bootstrap de escuelas fueron mayores que los del modelo para los coeficientes de la implementación (0,258 frente a 0,166 sobre el nivel inicial en lengua originaria y 0,084 frente a 0,065 sobre la pendiente en castellano). La Figura 16-4 reúne la potencia estimada por Monte Carlo frente a la muestra realizada (panel A), la retención por región frente a la prevista (panel B), los predictores de la permanencia (panel C) y la correlación intraclase de cada puntuación frente a la de planificación (panel D).

16.5 DISCUSIÓN

El hallazgo más desconcertante no es que el castellano crezca más deprisa que la lengua originaria entre tercero y quinto grado, sino que la lengua que menos avanza sea la que antecede al avance de la otra. Con pendientes medias de 1,59 y 2,28 puntos por año, una lectura de suma cero predeciría que el tiempo escolar dedicado a la lengua originaria frena el castellano; los datos señalan lo contrario. Dentro de cada estudiante, una comprensión en lengua originaria superior a su propio promedio precedió a una superior en castellano un año después, y el nivel inicial en lengua originaria predijo la velocidad posterior de aprendizaje en castellano. El efecto inverso también fue positivo, aunque menor, de modo que la relación parece recíproca con un sesgo direccional y no una vía de sentido único. Esta configuración es la que cabe esperar si ambas lenguas se apoyan en una competencia subyacente común, como propone la hipótesis de la interdependencia (Cummins, 1979), y resulta difícil de conciliar con la idea de que las dos alfabetizaciones compiten por un recurso escaso. La transferencia, además, persistió tras aislar interceptos aleatorios que absorbieron entre el 63,6 % y el 71,9 % de la varianza, lo que sugiere que no es un mero artefacto de diferencias estables entre estudiantes.

En contraste con parte de la literatura longitudinal, el patrón observado no se reduce a relaciones internas de cada lengua. Giguere y Hoff (2024)

encontraron que el vocabulario a los cinco años predecía el desarrollo de la comprensión lectora dentro de cada lengua, y no entre lenguas, en niños bilingües español-inglés; sin embargo, su predictor era léxico y el nuestro es la comprensión misma, un proceso de orden superior. Esa distinción es coherente con el hallazgo de Xie y Yeung (2026), quienes observaron transferencia de la comprensión lectora y de la composición morfológica entre el chino y el inglés, pero no del vocabulario ni de la lectura de palabras. Burchell et al. (2023) muestran, en la misma línea, que dentro del francés la relación entre conciencia sintáctica y comprensión estuvo mediada por la lectura de palabras y el vocabulario de esa lengua, es decir, por componentes específicos de cada lengua; si estos absorben lo que es propio de cada código, lo que cruza de una lengua a otra podría ser la construcción del modelo del texto. La evidencia sobre la comprensión en segunda lengua en educación bilingüe temprana apunta igualmente a una arquitectura por capas, donde la calidad léxica y la decodificación en la segunda lengua sostienen la comprensión en esa lengua (Kwakkel et al., 2024), y la metarrevisión de la evidencia correlacional ya había advertido que las correlaciones entre lenguas varían mucho según el componente considerado (Melby y Lervåg, 2011). Nuestra medida, que promedia subescalas literal, inferencial y crítico-reflexiva, no descompone esas capas, de modo que no puede decidir qué proceso se transfiere. Tampoco puede descartarse que la asimetría cambie con el tiempo: Shen y Huang (2026) describen una dominancia cambiante de los predictores a lo largo de la primaria, y la ventana de tercero a quinto grado podría capturar solo un tramo de esa dinámica.

La distancia tipológica ofrece una segunda vía de lectura. Zhang et al. (2025) atribuyen las asociaciones más fuertes de conciencia morfológica entre el español y el inglés a similitudes de vocabulario y morfología léxica entre ambas lenguas. Que aquí la transferencia aparezca entre lenguas aglutinantes y el castellano, sin tal similitud, sugiere que no depende de ella, aunque el diseño no compara pares de lenguas de distinta distancia y la interpretación es provisional. La magnitud de la transferencia también depende de factores individuales y de la exposición: la autoeficacia en la

primera lengua facilita la transferencia de la comprensión a la segunda (Liu et al., 2024) y las diferencias individuales en el desarrollo de dos lenguas se vinculan con la exposición a cada una (Paradis, 2023). En esta cohorte, el uso de la lengua originaria en el hogar se asoció con un nivel inicial mayor en esa lengua y ligeramente menor en castellano, sin relación con el crecimiento, un patrón compatible con un reparto de la exposición pero no con un costo para la trayectoria posterior. La autoeficacia no se midió.

Desde una perspectiva metodológica, el modelo elegido modificó la lectura de los datos. El modelo tradicional sobrestimó la estabilidad y produjo efectos cruzados de otra magnitud, con un ajuste mucho peor, lo que confirma en un dominio nuevo la advertencia de que la estabilidad propia del rasgo no queda recogida por las relaciones autorregresivas (Hamaker et al., 2015). Grimm et al. (2021) hallaron que la fuerza de las asociaciones rezagadas entre lectura y matemáticas dependía del modelo, y Mund et al. (2021) muestran que el tamaño y la interpretación de los parámetros intrapersona cambian con las covariables estables y variables en el tiempo; nuestras estimaciones deben leerse con esa cautela. La correlación de 0,54 entre los interceptos aleatorios de ambas lenguas se asemeja a la fuerte correlación interpersona que Psyridou et al. (2025) encontraron entre fluidez lectora y aritmética, y recuerda que quien domina una lengua tiende a destacar también en la otra. En términos de magnitud, el efecto de la lengua originaria sobre el castellano duplica el percentil 75 de la distribución de referencia, pero esa comparación se apoya en estimaciones estandarizadas de un intervalo y no elimina la imprecisión del modelo alternativo. La curva latente con residuos estructurados reprodujo el orden de los efectos con intervalos más amplios y una diferencia menor entre direcciones, y en las escuelas aimaras el efecto inverso rozó la significación; la asimetría direccional es, por tanto, menos firme que la existencia de transferencia bidireccional.

Las asociaciones con la implementación admiten una lectura de política. Una desviación estándar más de implementación se vinculó con 0,10 DE adicionales de crecimiento en castellano en dos años, un valor cercano al tramo inferior de los efectos de programas sobre la lectura en el Perú

rural, 0,075 DE para el acompañamiento pedagógico (Castro et al., 2025) y de 0,137 a 0,330 DE para el aprendizaje del quechua con apoyo móvil (Näslund et al., 2026), y muy por debajo de los efectos de la educación bilingüe de transición con materiales en lengua materna, de 0,41 y 0,75 DE (Leighton, 2022). Esa posición es plausible para una variación de intensidad dentro de escuelas que ya aplican el modelo y no para la comparación entre aplicarlo o no. El efecto indirecto por el nivel inicial en lengua originaria sugiere, además, que el beneficio para el castellano pasa en parte por la alfabetización previa en la lengua del niño y no es un efecto paralelo.

La calidad de la implementación es el eslabón donde la región muestra su mayor fragilidad. Las etnografías y los análisis de formación docente en Chile, Guatemala y Ecuador describen la distancia entre la norma intercultural y el aula: un educador tradicional ante un mandato intercultural (Aguayo et al., 2023), una formación docente que ensaya enfoques alternativos de descolonización curricular (Cardani, 2024) y registros lingüísticos y usos del kichwa divergentes entre escuelas (Limerick, 2023). En el Perú, los docentes de comunidades nativas describen las competencias interculturales bilingües como algo que excede las habilidades lingüísticas e incluye el compromiso con la comunidad (Chávez et al., 2026). Nuestro índice abarca solo tiempo de instrucción, materiales y dominio docente, con una fiabilidad moderada; dimensiones como el vínculo comunitario quedan fuera, y su omisión probablemente atenúa las asociaciones estimadas. Las asociaciones observadas sugieren que el tiempo efectivo en lengua originaria, los materiales y el dominio docente son palancas sobre las que la formación inicial y en servicio puede actuar.

Varias limitaciones acotan estas lecturas. El diseño es observacional: la implementación no se asignó y puede correlacionarse con características no medidas de escuelas o comunidades, de modo que las asociaciones no deben leerse como efectos causales; un diseño con asignación de apoyos a la implementación permitiría estimarlos. Las escuelas se incorporaron por conveniencia, y la generalización a todas las escuelas EIB de la región

requiere replicación. El índice de implementación se midió una sola vez, al inicio. Con tres olas, el modelo con residuos estructurados estimó los efectos cruzados con poca precisión, y separar el crecimiento individual de la dinámica intrapersona requiere más olas. El desgaste dependió de la comprensión en castellano y del nivel socioeconómico; la estimación con información completa es apropiada bajo ese mecanismo, pero no corrige un desgaste que dependa de lo no observado. Finalmente, la variable de escuela se analizó en modelos de un nivel con bootstrap de escuelas; un modelo de ecuaciones estructurales multinivel permitiría separar la variación de las trayectorias entre y dentro de las escuelas. Replicar el diseño con más olas, con mediciones repetidas de la implementación y con pares de lenguas de distinta distancia tipológica permitiría contrastar las interpretaciones propuestas.

16.6 CONCLUSIONES

En una cohorte de estudiantes quechuahablantes y aimarahablantes seguida de tercero a quinto grado, la comprensión lectora creció en ambas lenguas y se midió con una escala invariante a lo largo del tiempo. Una implementación más intensa de la educación intercultural bilingüe se asoció con un nivel inicial mayor en lengua originaria y con un crecimiento más rápido en castellano, en parte por la vía de ese nivel inicial. Dentro de cada estudiante, el progreso en lengua originaria antecedió al progreso en castellano más que a la inversa, y este resultado se mantuvo en las especificaciones de robustez, aunque con menor precisión en el subgrupo aimara y en el modelo con residuos estructurados.

Estos resultados tienen tres consecuencias. En lo teórico, son consistentes con la hipótesis de la interdependencia y contradicen, para este contexto, la idea de que la alfabetización en lengua originaria compite con la del castellano; la evidencia procede de una prueba intrapersona entre lenguas tipológicamente distantes y en alfabetización inicial con recursos escasos. En lo metodológico, mostraron que separar las diferencias estables de la dinámica dentro de cada estudiante modifica la magnitud de los efectos cruzados, de modo que los estudios de transferencia deberían informar

más de un modelo. En lo práctico, la calidad con que se implementa la política, medida por el tiempo de instrucción en lengua originaria, los materiales y el dominio docente de la lengua, es una variable observable que los sistemas pueden vigilar y fortalecer.

Dado su carácter observacional y muestral por conveniencia, las conclusiones describen asociaciones en las escuelas analizadas y no efectos causales de la política. Para cualquier sistema que enseñe a leer en una lengua minorizada y en una lengua dominante, el mensaje es que la alfabetización en la lengua del niño no resta: antecede y sostiene el aprendizaje de la segunda lengua, y la calidad de su implementación en la escuela merece medirse y fortalecerse.

16.7 REFERENCIAS BIBLIOGRÁFICAS

Aguayo, T., Contreras, S., & Giebeler, C. (2023). Intercultural bilingual education in Chile: The challenge of a traditional educator. *Journal of Higher Education Theory and Practice, 23*(18). https://doi.org/10.33423/jhetp.v23i18.6651

Burchell, D., Hipfner, K., Deacon, S., Koh, P., & Chen, X. (2023). Syntactic awareness and reading comprehension in emergent bilingual children. *Languages, 8*(1), 62. https://doi.org/10.3390/languages8010062

Cardani, G. (2024). Teacher training in intercultural bilingual education in Guatemala: A tool for critical decolonization? *Oxímora. Revista Internacional de Ética y Política*, (24), 134-159. https://doi.org/10.1344/oximora.24.2024.43802

Castro, J., Glewwe, P., Heredia, A., Majerowicz, S., & Montero, R. (2025). Can teaching be taught? Improving teachers' pedagogical skills at

scale in rural Peru. *Quantitative Economics, 16*(1), 185-233. https://doi.org/10.3982/qe2079

Chávez, R., Ruiz, Y., Carranza, R., Campos, J., Guevara, R., Navas, C., & Poémape, C. (2026). Intercultural bilingual education in teacher education within native communities of the Amazonas region. *Studies in Higher Education*, 1-20. https://doi.org/10.1080/03075079.2026.2679171

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251. https://doi.org/10.3102/00346543049002222

Curran, P., Howard, A., Bainter, S., Lane, S., & McGinley, J. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879-894. https://doi.org/10.1037/a0035297

Flores, H., Gamarra, J., Mamani, C., & Vilcarromero, L. (2023). Intercultural bilingual education (EIB) in Peru. *Research in Subject-matter Teaching and Learning, 6*(1), 84-95. https://doi.org/10.2478/ristal-2023-0006

Giguere, D., & Hoff, E. (2024). Bilingual children's vocabulary skills at 5 years predict reading comprehension development within, not across, languages. *International Journal of Bilingual Education and Bilingualism, 27*(2), 240-252. https://doi.org/10.1080/13670050.2023.2166397

Glazerman, S., Campuzano, L., & Murray, N. (2025). Education experiments in Latin America: Empirical evidence to guide evaluation design. *Evaluation Review*, *49*(1), 115-146. https://doi.org/10.1177/0193841x241241354

Grimm, K., Helm, J., Rodgers, D., & O'Rourke, H. (2021). Analyzing cross-lag effects: A comparison of different cross-lag modeling approaches. *New Directions for Child and Adolescent Development*, *2021*(175), 11-33. https://doi.org/10.1002/cad.20401

Groskurth, K., Bluemke, M., & Lechner, C. (2023). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods*, *56*(4), 3891-3914. https://doi.org/10.3758/s13428-023-02193-3

Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. https://doi.org/10.1037/a0038889

Kvietok, F. (2023). Migrant bilingual youth, family, and school language policy: Ethnographic insights for urban Quechua education. *International Journal of the Sociology of Language*, *2023*(280), 143-166. https://doi.org/10.1515/ijsl-2022-0016

Kwakkel, H., Droop, M., Verhoeven, L., & Segers, E. (2024). How lexical quality predicts L2 reading comprehension in early bilingual education. *International Journal of Bilingual Education and Bilingualism*, *27*(7), 993-1007. https://doi.org/10.1080/13670050.2024.2317860

Lasker, J. (2024). Measurement invariance testing works. *Applied Psychological Measurement, 48*(6), 257-275. https://doi.org/10.1177/01466216241261708

Leighton, M. (2022). Mother tongue reading materials as a bridge to literacy. *Economics of Education Review, 91,* 102312. https://doi.org/10.1016/j.econedurev.2022.102312

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research, 110,* 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Limerick, N. (2023). Linguistic registers and citizenship education: Divergent approaches to content, instruction, Kichwa use, and state relationships in Ecuador's intercultural bilingual education. *American Educational Research Journal, 60*(2), 219-256. https://doi.org/10.3102/00028312231152584

Liu, Y., Cheong, C., Ng, R., & Tse, S. (2024). The role of L1 self-efficacy in L2 reading comprehension: An exploration of L1–L2 cross-linguistic transfer. *International Journal of Bilingual Education and Bilingualism, 27*(7), 883-897. https://doi.org/10.1080/13670050.2023.2290478

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling.

Psychological Methods, 1(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Melby, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading, 34*(1), 114-135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Moreano, G., Zegarra, C., & Rivera, Y. (2026). New intercultural bilingual education teacher preparation curriculum: Opportunities and challenges to reduce educational inequities in Peru. *Education Policy Analysis Archives, 34.* https://doi.org/10.14507/epaa.34.9008

Moshagen, M., & Bader, M. (2023). semPower: General power analysis for structural equation models. *Behavior Research Methods, 56*(4), 2901-2922. https://doi.org/10.3758/s13428-023-02254-7

Mulder, J., & Hamaker, E. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 638-648. https://doi.org/10.1080/10705511.2020.1784738

Mund, M., Johnson, M., & Nestler, S. (2021). Changes in size and interpretation of parameter estimates in within-person models in the presence of time-invariant and time-varying covariates. *Frontiers in Psychology, 12,* 666928. https://doi.org/10.3389/fpsyg.2021.666928

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation*

Modeling: A Multidisciplinary Journal, *9*(4), 599-620. https://doi.org/10.1207/S15328007SEM0904_8

Näslund, E., Hernández, J., Santos, H., & Us, H. (2026). Mobile-assisted indigenous language learning: RCT evidence from a school-based intervention in Peru. *NABE Journal of Research and Practice*, 1-32. https://doi.org/10.1080/26390043.2026.2644295

Orth, U., Meier, L., Bühler, J., Dapp, L., Krauss, S., Messerli, D., & Robins, R. (2024). Effect size guidelines for cross-lagged effects. *Psychological Methods, 29*(2), 421-433. https://doi.org/10.1037/met0000499

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793-817. https://doi.org/10.1017/s0305000922000708

Paredes, R., Vilca, O., Arpi, R., Sucari, H., Huamani, A., Yapuchura, C., Gonzales, M., Quispe, C., Rodríguez, O., & Condori, J. (2026). Assessing the impact of socioeconomic and school factors on learning outcomes in Peru: A spatial perspective. *Frontiers in Education, 10,* 1600944. https://doi.org/10.3389/feduc.2025.1600944

Psyridou, M., Tolvanen, A., Koponen, T., Aunola, K., Lerkkanen, M., Poikkeus, A., & Torppa, M. (2025). Directional associations in reading and arithmetic fluency development across grades 1 to 9: A random intercept cross-lagged panel model. *Developmental Psychology*, *61*(6), 1196-1209. https://doi.org/10.1037/dev0001944

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2—An improved criterion for assessing discriminant validity in structural equation

modeling. *Industrial Management & Data Systems, 121*(12), 2637-2650. https://doi.org/10.1108/imds-02-2021-0082

Shen, Y., & Huang, B. (2026). Shifting dominance: Bilingual language skills in Spanish–English bilinguals' English reading comprehension over time. *Reading and Writing, 39*(8), 3131-3161. https://doi.org/10.1007/s11145-026-10762-z

Tang, D., & Tong, X. (2025). Evaluation of missing data analytical techniques in longitudinal research: Traditional and machine learning approaches. *Psychological Methods.* https://doi.org/10.1037/met0000765

Tarrillo, J., Mamani, L., Tocto, E., Iftikhar, H., Gonzales, R., & Soria, J. (2026). Exploring patterns in intercultural bilingual education in Peru using machine learning. *Frontiers in Education, 11*, 1832249. https://doi.org/10.3389/feduc.2026.1832249

Xie, Q., & Yeung, S. (2026). Do word-level skills show cross-linguistic transfer and predict bilingual reading? *Reading and Writing, 39*(8), 3275-3299. https://doi.org/10.1007/s11145-026-10767-8

Zhang, K., Sun, X., Flores, Z., Yu, C., Eggleston, R., Nickerson, N., Caruso, V., Tardif, T., & Kovelman, I. (2025). Cross-linguistic transfer in bilingual children's phonological and morphological awareness skills: A longitudinal perspective. *Bilingualism: Language and Cognition, 28*(2), 327-342. https://doi.org/10.1017/s1366728924000439

Zheng, B., & Valente, M. (2023). Evaluation of goodness-of-fit tests in random intercept cross-lagged panel model: Implications for small samples. *Structural Equation Modeling: A Multidisciplinary Journal,*

30(4), 604-617. https://doi.org/10.1080/10705511.2022.2149534

Zsögön, M. (2025). Between assimilation and recognition of Indigenous Peoples: The role of intercultural bilingual education in Latin America. *International Journal of Social Pedagogy*, *14*(1). https://doi.org/10.14324/111.444.ijsp.2025.v14.x.002

CAPÍTULO XVII

RETROALIMENTACIÓN AUTOMATIZADA CON INTELIGENCIA ARTIFICIAL GENERATIVA Y PRECISIÓN SINTÁCTICA EN LA ESCRITURA EN INGLÉS COMO LENGUA EXTRANJERA: ENSAYO ALEATORIZADO DE TRES BRAZOS EN UNIVERSIDADES PERUANAS

Diego Alejo Morales Aguero

Universidad nacional santiago antúnez de mayolo

Resumen

La retroalimentación generada por modelos de lenguaje se ha incorporado con rapidez a la enseñanza de la escritura en inglés como lengua extranjera, pero su efecto sobre la precisión sintáctica se ha evaluado sobre todo con diseños de un solo grupo y mediciones inmediatas. Este ensayo aleatorizado de tres brazos comparó la retroalimentación con inteligencia artificial generativa (RIA), la retroalimentación correctiva docente (RD) y un control centrado en el contenido en 432 universitarios de nivel B1 de tres universidades peruanas, con pretest, postest y postest diferido a las ocho semanas. La precisión se midió en 1240 textos con la proporción de cláusulas sin error (CSE/C), la de unidades T sin error y los errores por 100 palabras, cuya estructura fue invariante entre mediciones. En modelos lineales mixtos, la RIA superó al control en el postest en CSE/C (g = 0.56, IC 95 % [0.33, 0.79]) y la RD también lo hizo (g = 0.28); la ventaja de la RIA sobre la RD no resistió el ajuste por comparaciones múltiples. En el diferido, la ventaja de la RIA se mantuvo (g = 0.42, IC 95 % [0.19, 0.65]) y la de la RD dejó de ser significativa. El compromiso con la retroalimentación transmitió el 31 % del efecto total de la RIA, que fue mayor en los estudiantes con menor competencia inicial, y la reducción de errores se concentró en tiempos verbales, concordancia y artículos. La retroalimentación generativa se perfila como un complemento de magnitud al menos docente, duradero y dependiente del compromiso del estudiante.

Palabras clave: retroalimentación correctiva escrita; inteligencia artificial generativa; precisión sintáctica; escritura en lengua extranjera; compromiso con la retroalimentación; modelos lineales mixtos.

Abstract

Feedback generated by large language models has rapidly entered English as a foreign language writing instruction, yet its effect on syntactic accuracy has been evaluated mostly with single-group designs and immediate measurements.

This three-arm randomized trial compared generative artificial intelligence feedback (AIF), teacher corrective feedback (TF) and a content-focused control in 432 B1-level undergraduates from three Peruvian universities, with a pretest, a posttest and a delayed posttest eight weeks later. Accuracy was measured in 1240 texts through the proportion of error-free clauses (EFC/C), the proportion of error-free T-units and errors per 100 words, whose structure was invariant across occasions. In linear mixed models, AIF outperformed the control at posttest on EFC/C (g = 0.56, 95% CI [0.33, 0.79]) and TF did so as well (g = 0.28); the advantage of AIF over TF did not survive adjustment for multiple comparisons. At the delayed posttest, the advantage of AIF was maintained (g = 0.42, 95% CI [0.19, 0.65]), whereas that of TF was no longer significant. Engagement with feedback carried 31% of the total effect of AIF, which was larger for learners with lower initial proficiency, and error reduction concentrated on verb tense, agreement and articles. Generative feedback emerges as a complement of at least teacher-level magnitude that is durable and depends on learner engagement.

Keywords: *written corrective feedback; generative artificial intelligence; syntactic accuracy; foreign language writing; engagement with feedback; linear mixed models.*

17.1 INTRODUCCIÓN

La precisión gramatical de la escritura en una lengua extranjera es uno de los indicadores más estables del dominio lingüístico y uno de los más costosos de desarrollar en la enseñanza, porque exige retroalimentación individual, oportuna y repetida sobre textos completos (Barrot y Agdeppa, 2021; Phuoc y Barrot, 2022). La síntesis de la investigación sobre retroalimentación correctiva escrita ha resuelto en buena medida la pregunta sobre su eficacia general: un metaanálisis de 21 estudios mostró que mejora la precisión gramatical y que su efecto depende de la competencia del aprendiz, del contexto y del género textual (Kang y Han, 2015), y una actualización bayesiana con 52 estudios con grupo control encontró efectos moderados que se mantienen en el tiempo y magnitudes similares entre los tipos directo, indirecto y metalingüístico (Brown et

al., 2026). El problema práctico es otro: en las aulas universitarias de lengua extranjera, con grupos numerosos y pocas horas de docencia, la retroalimentación correctiva individual sobre cada borrador es difícil de sostener.

Las herramientas de evaluación automatizada de la escritura ofrecieron una primera respuesta. Los metaanálisis sitúan su efecto sobre el desempeño escrito en un rango moderado a grande: g = 0.59 entre grupos en 24 estudios con estudiantes de inglés como lengua extranjera o segunda lengua (Ngo et al., 2024), g = 0.55 en un metaanálisis multinivel de retroalimentación automatizada (Fleckenstein et al., 2023) y g = 0.861 sobre la calidad de la escritura en 26 estudios (Zhai y Ma, 2023), con beneficios también sobre la ansiedad y la motivación (Huang et al., 2024). Sin embargo, la misma literatura advierte que estas herramientas son menos eficaces para la gramática que para el vocabulario (Ngo et al., 2024), y los estudios con retroalimentación correctiva automatizada documentan mejoras de precisión junto con sobrecorrección, sobrecarga cognitiva y explicaciones metalingüísticas limitadas (Barrot, 2023; Guo et al., 2022; Fan, 2023; Tan et al., 2023; Zhai, 2025).

Los modelos de lenguaje generativos cambian la naturaleza de esa retroalimentación: pueden explicar un error, proponer una regla y adaptar el registro de la explicación al nivel del estudiante, en lugar de limitarse a marcar y sustituir (Kang y Yi, 2023; Banihashem et al., 2024; Abudalfa y Barrot, 2026; Liu et al., 2025). La investigación empírica sobre su efecto crece con rapidez, pero presenta tres limitaciones. Primero, buena parte de los estudios usa diseños de un grupo o cuasiexperimentales con medidas globales de calidad (Polakova y Ivenz, 2024; Asadi et al., 2025), y los que se centran en la precisión encuentran mejoras parciales con muchos errores persistentes (Pretorius y Thewissen, 2026; Fujisawa y Shintani, 2025). Segundo, la comparación relevante para la práctica —la retroalimentación generativa frente a la docente— rara vez se hace con aleatorización y con una condición control a la vez (Khojasteh et al., 2025; Wang et al., 2026; Li y Collins, 2026). Tercero, casi todos los estudios miden solo el efecto inmediato, aunque la durabilidad es precisamente lo que distingue el

aprendizaje de la corrección de un texto (Truscott, 2023; Kim y Chon, 2025).

A estas limitaciones se suma una pregunta sobre el mecanismo. La retroalimentación no actúa por sí sola: su efecto depende de que el estudiante la procese, la use para revisar y la valore, es decir, de su compromiso cognitivo, conductual y afectivo con ella (Ranalli, 2021; Zhang y Hyland, 2022). Los estudios con retroalimentación generativa describen un compromiso alto pero heterogéneo (Yeung, 2025; Ma et al., 2026; Kailin y Saeed, 2026) que cambia con el tiempo (Xiaosa y Ping, 2025) y que depende de la alfabetización digital y crítica del estudiante (Zhang y Hyland, 2025; Darvin, 2025; Wang y Wang, 2025), y un estudio con 174 universitarios encontró que la retroalimentación de ChatGPT aumentó a la vez el compromiso afectivo y conductual y el desempeño escrito (Teng y Huang, 2026). Falta, sin embargo, una estimación del efecto indirecto de la retroalimentación generativa sobre la precisión a través del compromiso, con un modelo que distinga ese camino del efecto directo y que controle la precisión inicial.

El contexto peruano hace la pregunta especialmente pertinente. En las universidades peruanas el inglés se enseña en cursos de pocas horas con grupos numerosos, y la investigación local ha documentado la relación entre la autoeficacia y las estrategias de aprendizaje del inglés (Saez et al., 2025) y el uso de entornos virtuales y recursos digitales en su enseñanza (La Torre, 2021; Gutiérrez, 2024), pero no ha evaluado con diseños experimentales el efecto de la retroalimentación generativa sobre la escritura. Al mismo tiempo, el acceso de los estudiantes a estas herramientas ya es generalizado (Teng, 2024; Hwang et al., 2025), de modo que la decisión institucional no es si se usan, sino cómo integrarlas para que produzcan aprendizaje y no solo textos corregidos.

Este estudio responde a esas tres limitaciones con un ensayo aleatorizado de tres brazos, una medición diferida y un modelo del mecanismo, y aporta evidencia que trasciende el caso peruano: estima, con medidas de precisión sintáctica de fiabilidad conocida e invariantes entre mediciones,

si la retroalimentación generativa iguala a la docente, si su efecto perdura, por qué camino actúa y para quién es más útil. El aporte es triple: empírico, porque ofrece una comparación aleatorizada con control activo y medición diferida en educación superior latinoamericana; metodológico, porque integra modelos lineales mixtos con la estructura anidada de las aulas, invarianza longitudinal de la medida y un modelo de ecuaciones estructurales del mecanismo; y teórico, porque contrasta la hipótesis de que la retroalimentación generativa actúa a través del compromiso del estudiante y no solo por la cantidad de correcciones que ofrece.

17.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

17.2.1 La precisión sintáctica y su medición

En el marco de complejidad, precisión y fluidez, la precisión es el grado en que la producción se ajusta a las normas de la lengua meta, y en la escritura se mide con índices basados en unidades sintácticas —proporción de cláusulas o de unidades T sin error— y con tasas de error por extensión del texto (Barrot y Agdeppa, 2021). Estos índices discriminan entre niveles de competencia y cambian con la instrucción, aunque su comportamiento depende del nivel y de la lengua materna (Phuoc y Barrot, 2022). Medirlos en el mismo texto y combinarlos en un factor común permite separar el cambio en la precisión del ruido propio de cada índice, a condición de que la medida sea invariante entre mediciones (Luong y Flake, 2023; Mackinnon et al., 2022).

17.2.2 Retroalimentación correctiva docente, automatizada y generativa

La retroalimentación correctiva escrita favorece la precisión porque dirige la atención del aprendiz hacia la distancia entre su producción y la forma meta y, cuando incluye una explicación, hacia la regla que la gobierna (Kang y Han, 2015; Brown et al., 2026). La retroalimentación automatizada amplía la frecuencia y la inmediatez de ese aviso, pero su explicación suele ser breve y su precisión desigual entre categorías de error (Barrot, 2023; Guo et al., 2022). La retroalimentación generativa añade explicaciones

adaptadas y diálogo, y por eso puede acercarse a la docente en la calidad de la explicación (Banihashem et al., 2024; Wang et al., 2026); su riesgo es la reescritura completa del texto, que sustituye el procesamiento del error por la aceptación de una versión corregida, de modo que su efecto pedagógico depende de las salvaguardas con que se integra (Pretorius y Thewissen, 2026; Lu et al., 2026). Las categorías de error gobernadas por reglas —concordancia, artículos, tiempos verbales— son más sensibles a la retroalimentación que las que dependen de la estructura de la oración (Mujtaba et al., 2021; Peng, 2024).

17.2.3 El compromiso con la retroalimentación como mecanismo

El compromiso con la retroalimentación integra un componente cognitivo (atender y comprender la corrección), uno conductual (revisar y aplicar) y uno afectivo (valorar la retroalimentación y responder a ella), y su nivel explica por qué la misma retroalimentación produce aprendizajes distintos (Zhang y Hyland, 2022; Ranalli, 2021). Si la retroalimentación generativa mejora la precisión en parte porque provoca más procesamiento y más revisión, su efecto debe transmitirse a través del compromiso; si actúa solo por la cantidad de correcciones, el compromiso no debe mediar su efecto (Teng y Huang, 2026; Yeung, 2025).

17.2.4 Hipótesis

- H1. La RIA mejora la precisión sintáctica (CSE/C) en el postest más que la condición control.
- H2. La RD mejora la precisión sintáctica (CSE/C) en el postest más que la condición control.
- H3. La ventaja de la RIA sobre el control se mantiene en el postest diferido.
- H4. El compromiso con la retroalimentación media el efecto de la RIA sobre la precisión posterior, controlando la precisión inicial.

- H5. El efecto de la RIA es mayor en los estudiantes con menor competencia inicial.

Se formuló además una pregunta sin dirección prevista: ¿difiere el efecto de la RIA del de la RD en el postest y en el postest diferido? Las hipótesis H1 y H2 forman la familia confirmatoria primaria.

17.3 MATERIALES Y MÉTODOS

Se realizó un ensayo aleatorizado de tres brazos con pretest, postest tras diez semanas de intervención y postest diferido ocho semanas después, entre abril y agosto de 2025, en cursos obligatorios de inglés intermedio de tres universidades de Lima, Arequipa y Trujillo (24 secciones de 18 estudiantes). Dentro de cada sección, que actuó como estrato, los estudiantes se asignaron a las condiciones con bloques permutados generados por computadora antes del pretest, por lo que las tres condiciones compartieron docente, horario y temario. Participaron estudiantes de primer y segundo año con nivel B1 en la prueba de ubicación institucional; la selección de universidades y secciones fue no probabilística.

El tamaño muestral se fijó a priori en el mayor de tres requisitos. Para la familia confirmatoria primaria (RIA y RD frente al control en el postest, con ajuste de Holm para dos contrastes) se adoptó d = 0.50, inferior a los efectos de la retroalimentación automatizada sobre la escritura (g = 0.55 a 0.59; Fleckenstein et al., 2023; Ngo et al., 2024) y situado entre los umbrales pequeño y mediano de la investigación en segundas lenguas (Plonsky y Oswald, 2014), con ρ pretest-postest = .40 y potencia .80 (Vitta et al., 2022; Nicklin y Vitta, 2021): 69 estudiantes por condición. El modelo de ecuaciones estructurales (gl = 38) exigió 271 inscritos para distinguir un RMSEA de .05 de uno de .08 con potencia .80 (MacCallum et al., 1996), y la invarianza longitudinal (gl = 23), N = 388 con las tres mediciones; con una pérdida prevista del 10 %, se fijó N = 432 (144 por condición). La simulación del modelo lineal mixto (correlación intraclase de sección de .05, 300 réplicas) arrojó una potencia de .963 (Liu et al., 2024), y la del efecto indirecto, con intervalos Monte Carlo (Preacher y Selig, 2012; Xu

et al., 2024), fue .915 (1000 réplicas). El análisis por categoría de error fue secundario y exploratorio. Se perdieron 13 estudiantes en el postest y 43 (10.0 %) en el diferido; todos se analizaron según la condición asignada.

Durante diez semanas los estudiantes escribieron seis textos argumentativos de 250 a 300 palabras y entregaron una versión revisada tras cada retroalimentación; las condiciones difirieron solo en su fuente y foco. En la RIA, un asistente generativo con instrucción fija señalaba errores de cinco categorías (tiempo y aspecto verbal, concordancia sujeto-verbo, artículos y determinantes, preposiciones, orden de palabras), ofrecía una explicación metalingüística breve con un ejemplo y no reescribía el texto; los estudiantes entregaban el registro de la conversación (30.7 puntos de retroalimentación por texto; 72 % de revisiones adecuadas). En la RD, docentes capacitados durante seis horas ofrecieron retroalimentación indirecta codificada con explicación metalingüística sobre las mismas categorías (13.9 puntos; 80 %). En el control comentaron el contenido y la organización.

La precisión se midió en textos argumentativos de 35 minutos escritos a mano sin ayudas (260.3 palabras de media). Dos codificadores entrenados, ciegos a condición y medición, segmentaron cláusulas y unidades T y clasificaron 26 225 errores, de los que se derivaron CSE/C, UTSE/UT y E/100P; el 20 % de los textos (248) se codificó por duplicado y la fiabilidad se estimó con el CCI(2,1) (Koo y Li, 2016; Tabla 17-2). El compromiso se midió en la semana 12 con un cuestionario de doce ítems y tres subescalas (cognitiva, conductual y afectiva), y la competencia inicial con la prueba de ubicación de 60 ítems.

La equivalencia inicial se evaluó con ANOVA, χ^2 y la mayor g de Hedges entre pares. La estructura común de los índices se examinó con un análisis factorial confirmatorio longitudinal, contrastando la invarianza configural, métrica y escalar con $\Delta\chi^2$ y $\Delta\mathrm{CFI} \geq -.010$ (Chen, 2007; Luong y Flake, 2023; Mackinnon et al., 2022) y valorando el ajuste con Hu y Bentler (1999). El efecto se estimó para cada índice con un modelo lineal mixto de las tres mediciones, con condición, medición, sus interacciones,

competencia inicial estandarizada e interceptos aleatorios de sección y estudiante, por máxima verosimilitud restringida con todas las mediciones disponibles; los contrastes se expresaron como diferencias con IC 95 % y g de Hedges (sobre la DE del pretest) y se ajustaron por Holm dentro de cada índice. El cambio por categoría se analizó con ecuaciones de estimación generalizadas binomiales negativas y se expresó como razón de tasas de incidencia (RTI). El mecanismo se estimó con un modelo de ecuaciones estructurales con factores latentes de precisión previa, compromiso y precisión posterior, y el efecto indirecto con intervalos Monte Carlo de 20 000 extracciones. La moderación se contrastó con la interacción condición × competencia sobre el postest de CSE/C, con efectos condicionales a −1, 0 y +1 DE y región de Johnson-Neyman. La robustez se examinó con siete especificaciones alternativas, con $\alpha = .05$ bilateral.

17.4 RESULTADOS

17.4.1 Participantes y equivalencia inicial

Se aleatorizaron 432 estudiantes, 144 por condición; completaron el postest 419 (141, 137 y 141 en RIA, RD y C) y el postest diferido 389 (130, 129 y 130). Las tres condiciones no difirieron de forma significativa en las variables basales (Tabla 17-1); la mayor diferencia correspondió a CSE/C del pretest, y el modelo de análisis estima el cambio de cada condición desde su propio pretest.

Tabla 17-1

Características de los participantes y equivalencia inicial por condición (N = 432)

Variable	RIA	RD	C	Prueba	p	g máximo (valor absoluto)
	(n = 144)					
Edad (años)	20.3 (1.7)	20.5 (1.7)	20.0 (1.7)	$F(2, 429) = 2.64$	.073	0.27
Años de estudio del inglés	7.0 (2.0)	6.7 (2.2)	6.7 (2.0)	$F(2, 429) = 0.72$	.488	0.13
Prueba de ubicación (0-60)	34.7 (5.4)	35.0 (5.7)	34.9 (5.3)	$F(2, 429) = 0.16$	.856	0.06
Palabras por texto (pretest)	256.2 (42.5)	255.1 (39.8)	253.5 (39.7)	$F(2, 429) = 0.16$	.848	0.07
Cláusulas sin error / cláusulas (pretest)	0.544 (0.196)	0.598 (0.181)	0.562 (0.199)	$F(2, 429) = 2.88$	.057	0.28
Unidades T sin error / unidades T (pretest)	0.422 (0.212)	0.466 (0.213)	0.428 (0.215)	$F(2, 429) = 1.80$	.167	0.21
Errores por 100 palabras (pretest)	9.13 (3.85)	8.67 (3.58)	9.19 (4.22)	$F(2, 429) = 0.77$	.461	0.13
Mujeres, n (%)	89 (61.8)	82 (56.9)	77 (53.5)	$\chi^2(2) = 2.06$	.356	—

Nota: Valores M (DE), salvo la fila de mujeres, n (%). RIA = retroalimentación automatizada con inteligencia artificial; RD = retroalimentación correctiva docente; C = control. g máximo = mayor diferencia estandarizada de Hedges, en valor absoluto, entre los tres pares de condiciones. CSE/C = cláusulas sin error por cláusula; UTSE/UT = unidades T sin error por unidad T; E/100P = errores por cada 100 palabras.

17.4.2 Fiabilidad y estructura de la medida

La fiabilidad entre codificadores fue excelente en los tres índices y el acuerdo en la clasificación de los errores, casi perfecto; las tres subescalas de compromiso tuvieron una consistencia interna adecuada (Tabla 17-2). El factor común de precisión tuvo cargas estandarizadas entre .61 y .88 en las tres mediciones.

Tabla 17-2

Operacionalización de las medidas y fiabilidad

Medida	Definición operativa	Fuente	Fiabilidad [IC 95 %]	
CSE/C	Cláusulas sin ningún error morfosintáctico divididas por el total de cláusulas del texto	Codificación de los 1240 textos	CCI(2,1) =	.96 [.95, .97]
UTSE/UT	Unidades T sin error divididas por el total de unidades T			.96 [.94, .97]
E/100P	Errores morfosintácticos por cada 100 palabras (suma de las cinco categorías)			.94 [.92, .95]
Categoría de error	Tiempo y aspecto verbal, concordancia, artículos y determinantes, preposiciones, orden de palabras	Clasificación de 600 errores por dos codificadores	κ = .92 [.89, .94]	
Precisión sintáctica (factor latente)	Factor común de CSE/C, UTSE/UT y E/100P (invertido) en cada medición	Modelo de medida longitudinal (Tabla 4)	ω = .81 / .85 / .83 (pretest / postest / diferido)	
Compromiso cognitivo	Cuatro ítems sobre el procesamiento de la retroalimentación (escala 1-5)	Cuestionario de postest (n = 419)	ω = .80; α = .80	
Compromiso conductual	Cuatro ítems sobre las acciones de revisión (escala 1-5)		ω = .78; α = .78	
Compromiso afectivo	Cuatro ítems sobre la valoración de la retroalimentación (escala 1-5)		ω = .74; α = .73	

Nota: La fiabilidad de las tres medidas de precisión se estimó con el coeficiente de correlación intraclase de dos vías, efectos aleatorios y acuerdo absoluto, CCI(2,1), sobre 248 textos codificados de forma independiente por dos codificadores (20 % del total); κ = coeficiente kappa de Cohen; ω = coeficiente omega a partir de las cargas estandarizadas; α = coeficiente alfa de Cronbach.

17.4.3 Estadísticos descriptivos

La Tabla 17-3 presenta los descriptivos por condición y medición. En las dos condiciones con retroalimentación sobre la forma, la proporción de cláusulas sin error aumentó y la tasa de errores disminuyó entre el pretest y el postest, mientras que en el control los cambios fueron menores.

Tabla 17-3

Estadísticos descriptivos de la precisión sintáctica por condición y medición

Medida	Condición	Pretest	Postest	Postest diferido
		M (DE)		
CSE/C	RIA	0.544 (0.196)	0.647 (0.184)	0.634 (0.182)
	RD	0.598 (0.181)	0.646 (0.192)	0.642 (0.171)
	C	0.562 (0.199)	0.556 (0.178)	0.579 (0.183)
UTSE/UT	RIA	0.422 (0.212)	0.541 (0.208)	0.510 (0.216)
	RD	0.466 (0.213)	0.517 (0.221)	0.547 (0.189)
	C	0.428 (0.215)	0.442 (0.213)	0.449 (0.198)
E/100P	RIA	9.13 (3.85)	7.13 (2.86)	7.16 (3.07)
	RD	8.67 (3.58)	7.28 (2.99)	7.01 (2.61)
	C	9.19 (4.22)	9.44 (4.61)	8.76 (3.81)

Nota: n por condición: pretest 144, 144 y 144; postest 141, 137 y 141; postest diferido 130, 129 y 130 (RIA, RD y C). En CSE/C y UTSE/UT los valores más altos indican mayor precisión; en E/100P, los más bajos.

Tabla 17-4

Invarianza longitudinal del modelo de medida de la precisión sintáctica (n = 389)

Modelo	χ^2 (gl)	p	CFI	TLI	RMSEA [IC 90 %]	SRMR	Comparación: $\Delta\chi^2$ (Δgl); p; ΔCFI
Configural	11.95 (15)	.683	1.000	1.004	.000 [.000, .038]	.014	—
Métrica (cargas iguales)	14.43 (19)	.758	1.000	1.005	.000 [.000, .032]	.019	2.47 (4); p = .649; ΔCFI = 0.000
Escalar (cargas e interceptos iguales)	22.98 (23)	.462	1.000	1.000	.000 [.000, .042]	.024	8.55 (4); p = .073; ΔCFI = 0.000

Nota: Ajuste por máxima verosimilitud con estructura de medias de modelos anidados con tres factores, uno por medición, y covarianzas residuales del mismo indicador entre mediciones. Cargas estandarizadas del modelo escalar: CSE/C .85, .88 y .86; UTSE/UT .83, .85 y .82; E/100P invertido .61, .68 y .65 (pretest, postest y diferido); ω = .81, .85 y .83; varianza media extraída = .60, .65 y .62. Criterios de ajuste: CFI ≥ .95, RMSEA ≤ .06 y SRMR ≤ .08 (Hu y Bentler, 1999); invarianza aceptada si ΔCFI ≥ −.010 (Chen, 2007).

17.4.4 Invarianza longitudinal de la medida

El modelo configural ajustó bien a los datos, y ni la igualación de las cargas

17.4.5 Efecto de la retroalimentación

La interacción condición × medición fue significativa en los tres índices, y la competencia inicial predijo los tres (Tabla 17-5). La correlación intraclase de sección fue baja en CSE/C y UTSE/UT y algo mayor en E/100P.

Tabla 17-5

Modelos lineales mixtos de la precisión sintáctica: pruebas de Wald de los efectos fijos y componentes de varianzat

Efecto	CSE/C	UTSE/UT	E/100P
Multivariado (tres dimensiones): traza de Pillai = .097	19.28	3, 536	< .001
Condición	$\chi^2(2) = 6.52$; $p = .038$	$\chi^2(2) = 3.75$; $p = .153$	$\chi^2(2) = 1.85$; $p = .397$
Medición	$\chi^2(2) = 1.35$; $p = .508$	$\chi^2(2) = 1.15$; $p = .562$	$\chi^2(2) = 2.42$; $p = .298$
Condición × medición	$\chi^2(4) = 25.22$; $p < .001$	$\chi^2(4) = 20.45$; $p < .001$	$\chi^2(4) = 22.37$; $p < .001$
Competencia inicial (z)	$\chi^2(1) = 127.78$; $p < .001$	$\chi^2(1) = 103.30$; $p < .001$	$\chi^2(1) = 92.05$; $p < .001$
Varianza entre secciones (τ^2sec)	0.0008	0.0010	0.99
Varianza entre estudiantes (τ^2est)	0.0102	0.0145	2.24
Varianza residual (σ^2)	0.0180	0.0235	9.09
Correlación intraclase de sección	.029	.027	.081
R^2 marginal / condicional	.184 / .494	.164 / .496	.146 / .370
Observaciones (estudiantes; secciones)	1240 (432; 24)	1240 (432; 24)	1240 (432; 24)

Nota: Modelos estimados por máxima verosimilitud restringida con interceptos aleatorios de sección y de estudiante dentro de sección; la condición control y el pretest son las categorías de referencia. R^2 marginal = varianza explicada por los efectos fijos; R^2 condicional = por los efectos fijos y aleatorios.

17.4.6 Contrastes planificados

En el postest, la RIA superó al control en CSE/C (Δ = 0.108, IC 95 % [0.064, 0.152]; g = 0.56, p < .001) y la RD también (Δ = 0.054, IC 95 % [0.009, 0.098]; g = 0.28, p = .018); con el ajuste de Holm de la familia confirmatoria (dos contrastes), p < .001 y p = .018, de modo que H1 y H2 se sostuvieron (Tabla 17-6). El mismo patrón apareció en E/100P, con −2.22 errores por cada 100 palabras en la RIA respecto del control (g = −0.57, p Holm < .001) y −1.60 en la RD (g = −0.41, p Holm = .007). En UTSE/UT, la RIA superó al control (g = 0.48, p Holm < .001) y a la RD (g = 0.31, p Holm = .038), mientras que la RD no difirió del control (g = 0.17, p Holm = .309). En CSE/C, la diferencia entre la RIA y la RD (g = 0.28, IC 95 % [0.05, 0.51], p = .017) no superó el ajuste de los seis contrastes del índice (p Holm = .066).

En el postest diferido, la RIA mantuvo su ventaja sobre el control en CSE/C (g = 0.42, IC 95 % [0.19, 0.65], p Holm = .002), UTSE/UT (g = 0.35, p Holm = .025) y E/100P (g = −0.43, p Holm = .006), de modo que H3 se sostuvo. La ventaja de la RD dejó de ser significativa en CSE/C (g = 0.15, p = .211) y se mantuvo con menor magnitud en E/100P (g = −0.33, p Holm = .042). La diferencia entre la RIA y la RD en el diferido favoreció a la RIA en CSE/C (g = 0.27, IC 95 % [0.03, 0.50], p = .024), sin superar el ajuste (p Holm = .066). La Figura 17-1 resume las trayectorias, el cambio individual y los 18 contrastes.

Tabla 17-6

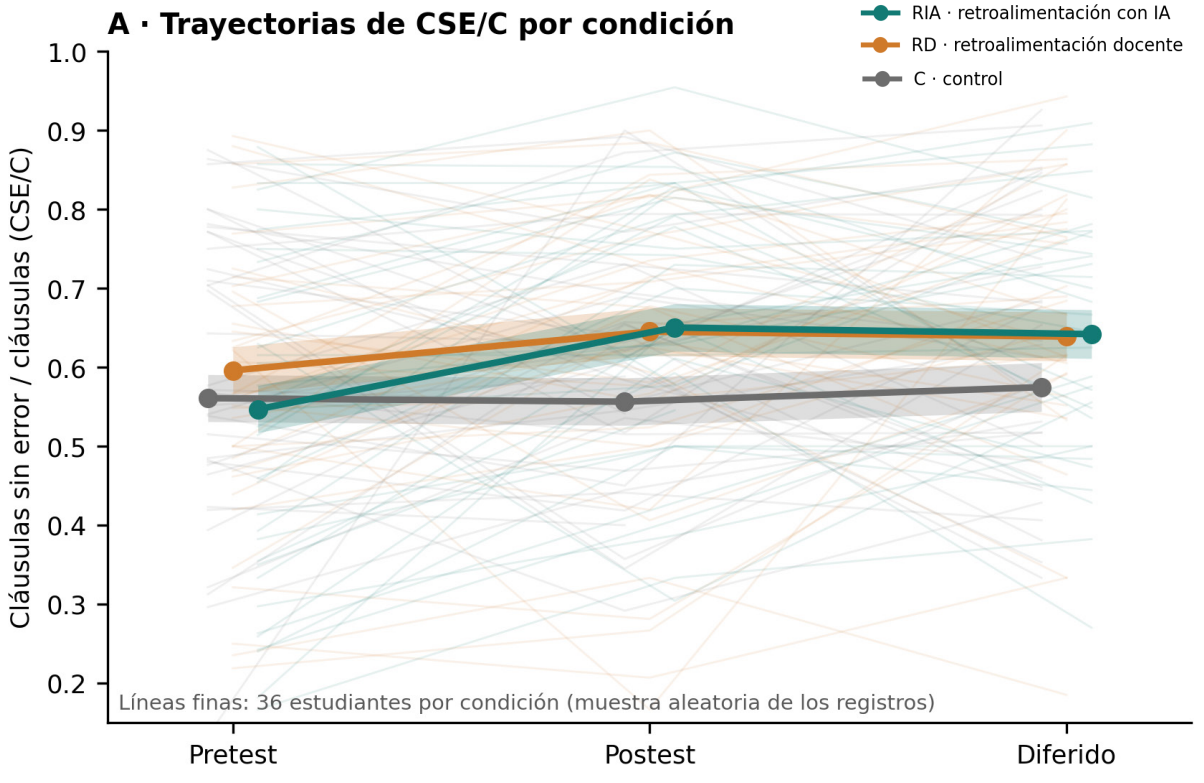
Contrastes planificados: diferencia entre condiciones en el cambio desde el pretest

Medida	Medición	Contraste	Δ [IC 95 %]	g [IC 95 %]	p absoluto	p Holm
CSE/C	Postest	RIA – C	0.108 [0.064, 0.152]	0.56 [0.33, 0.79]	< .001	< .001
		RD – C	0.054 [0.009, 0.098]	0.28 [0.05, 0.51]	.018	.066
		RIA – RD	0.054 [0.010, 0.099]	0.28 [0.05, 0.51]	.017	.066
	Diferido	RIA – C	0.081 [0.036, 0.127]	0.42 [0.19, 0.65]	< .001	.002
		RD – C	0.029 [–0.016, 0.074]	0.15 [–0.08, 0.38]	.211	.211
		RIA – RD	0.052 [0.007, 0.098]	0.27 [0.03, 0.50]	.024	.066
UTSE/UT	Postest	RIA – C	0.104 [0.053, 0.154]	0.48 [0.25, 0.72]	< .001	< .001
		RD – C	0.037 [–0.014, 0.088]	0.17 [–0.06, 0.41]	.154	.309
		RIA – RD	0.067 [0.016, 0.118]	0.31 [0.08, 0.55]	.010	.038
	Diferido	RIA – C	0.074 [0.022, 0.126]	0.35 [0.10, 0.59]	.005	.025
		RD – C	0.063 [0.011, 0.115]	0.29 [0.05, 0.54]	.018	.053
		RIA – RD	0.011 [–0.041, 0.063]	0.05 [–0.19, 0.29]	.672	.672
E/100P	Postest	RIA – C	–2.22 [–3.21, –1.23]	–0.57 [–0.82, –0.31]	< .001	< .001
		RD – C	–1.60 [–2.60, –0.60]	–0.41 [–0.67, –0.15]	.002	.007
		RIA – RD	–0.62 [–1.61, 0.38]	–0.16 [–0.41, 0.10]	.223	.447
	Diferido	RIA – C	–1.68 [–2.70, –0.66]	–0.43 [–0.69, –0.17]	.001	.006
		RD – C	–1.28 [–2.29, –0.26]	–0.33 [–0.59, –0.07]	.014	.042
		RIA – RD	–0.40 [–1.42, 0.61]	–0.10 [–0.36, 0.16]	.435	.447

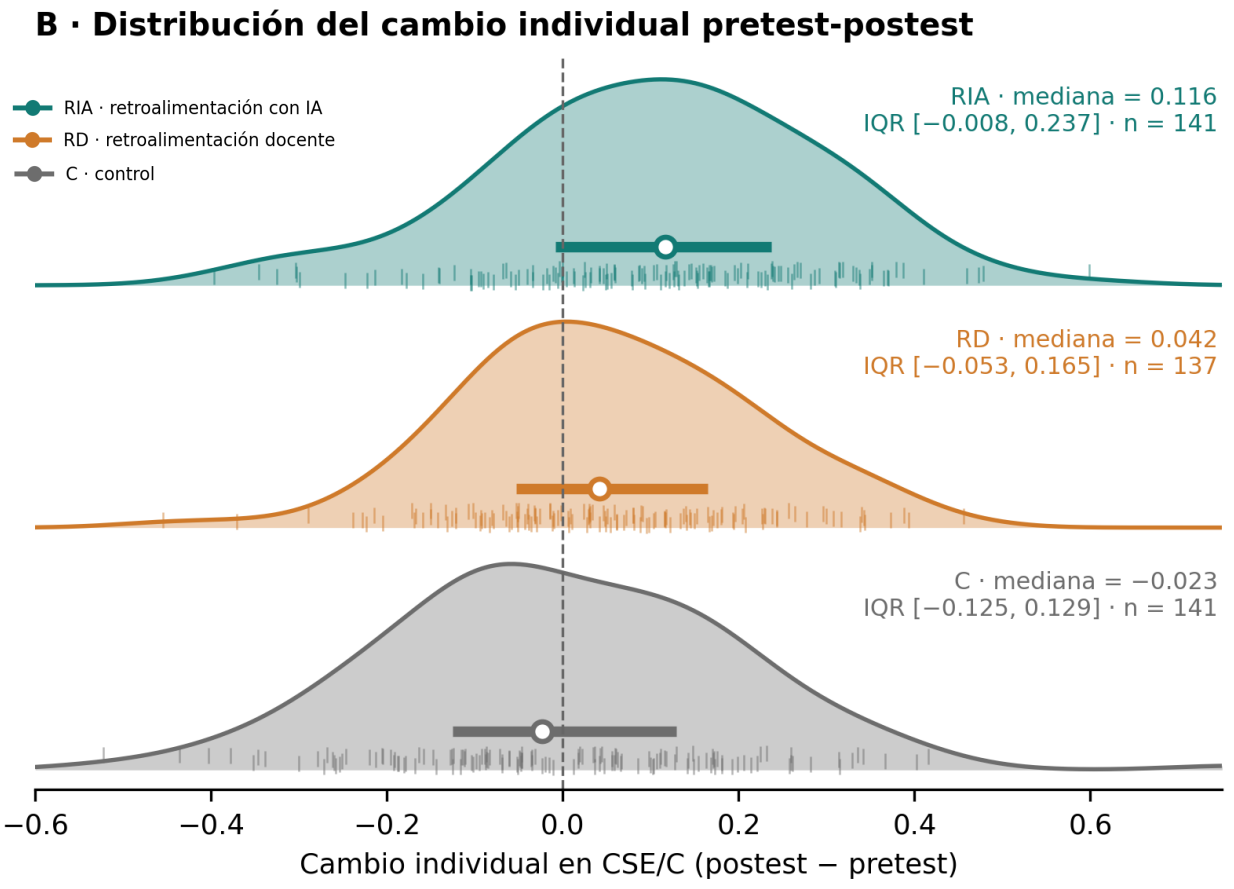
Nota: Δ = diferencia entre condiciones del cambio estimado desde el pretest (término de interacción del modelo lineal mixto de la Tabla 17-5); g = Δ dividida por la desviación típica del pretest con la corrección de Hedges. p (Holm) ajusta los seis contrastes de cada medida. En E/100P los valores negativos indican menos errores. Para la familia confirmatoria (RIA – C y RD – C en el postest de CSE/C), p (Holm, dos contrastes) < .001 y = .018.

Figura 17-1

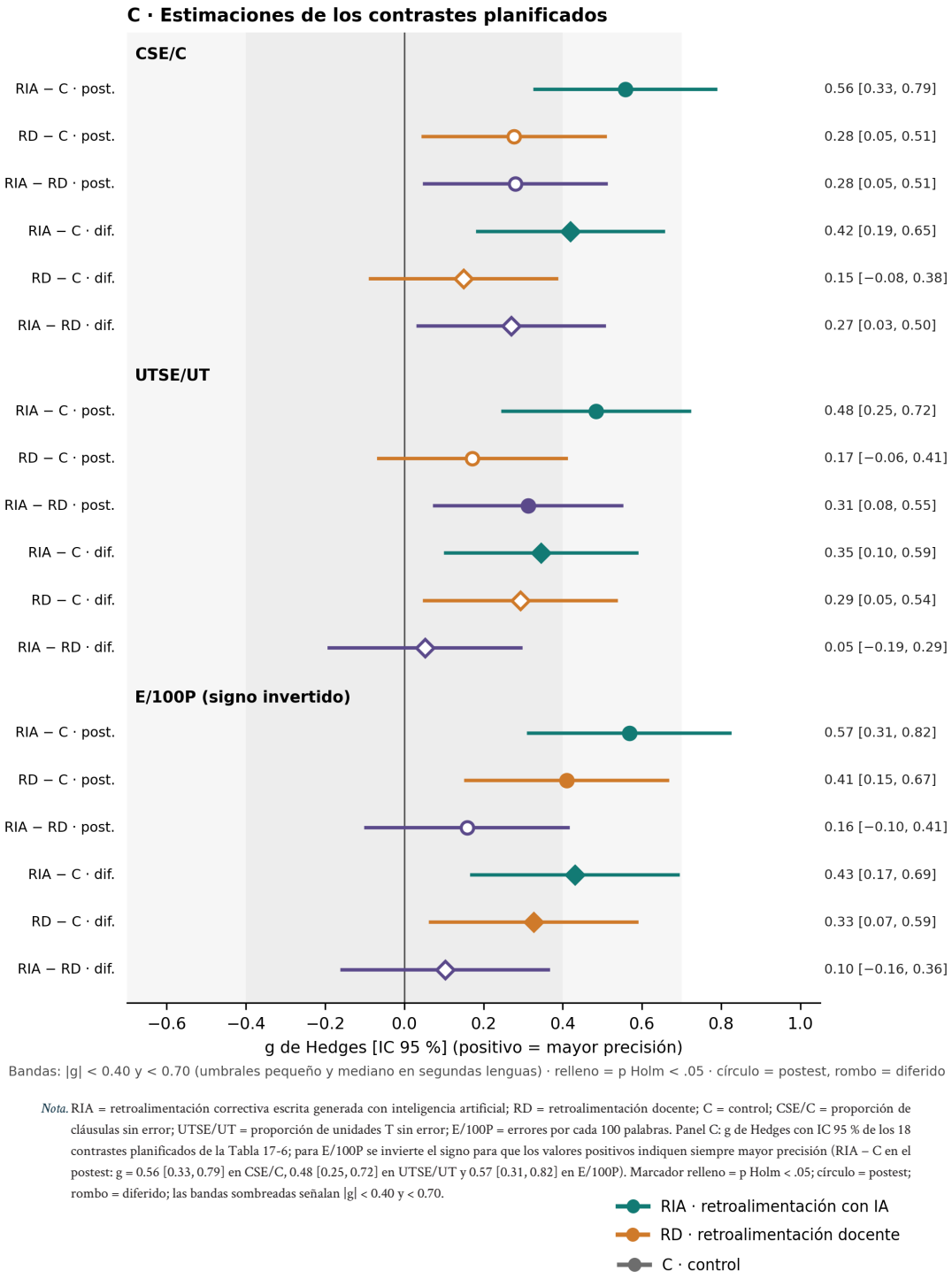
Precisión sintáctica por condición y medición



Nota. RIA = retroalimentación correctiva escrita generada con inteligencia artificial; RD = retroalimentación docente; C = control; CSE/C = proporción de cláusulas sin error; UTSE/UT = proporción de unidades T sin error; E/100P = errores por cada 100 palabras. Panel A: medias marginales estimadas de CSE/C con IC 95 % del modelo lineal mixto de la Tabla 17-5 (pretest n = 432; postest n = 419; diferido n = 389); las líneas finas representan las trayectorias observadas de 36 estudiantes por condición extraídos al azar. RIA: 0.547 → 0.650 → 0.642; RD: 0.596 → 0.645 → 0.639; C: 0.561 → 0.556 → 0.575.



Nota. RIA = retroalimentación correctiva escrita generada con inteligencia artificial; RD = retroalimentación docente; C = control; CSE/C = proporción de cláusulas sin error; UTSE/UT = proporción de unidades T sin error; E/100P = errores por cada 100 palabras. Panel B: densidad del cambio individual de CSE/C entre pretest y postest, con marcas de cada estudiante, mediana y rango intercuartílico (RIA: mediana 0.116, n = 141; RD: 0.042, n = 137; C: −0.023, n = 141).



Nota. RIA = retroalimentación correctiva escrita generada con inteligencia artificial; RD = retroalimentación docente; C = control; CSE/C = proporción de cláusulas sin error; UTSE/UT = proporción de unidades T sin error; E/100P = errores por cada 100 palabras. Panel C: g de Hedges con IC 95 % de los 18 contrastes planificados de la Tabla 17-6; para E/100P se invierte el signo para que los valores positivos indiquen siempre mayor precisión (RIA − C en el postest: g = 0.56 [0.33, 0.79] en CSE/C, 0.48 [0.25, 0.72] en UTSE/UT y 0.57 [0.31, 0.82] en E/100P). Marcador relleno = p Holm < .05; círculo = postest; rombo = diferido; las bandas sombreadas señalan |g| < 0.40 y < 0.70.

17.4.7 Categorías de error

Respecto del control, la RIA redujo en el postest la tasa de errores de tiempo y aspecto verbal en un 34 %, la de concordancia en un 30 % y la de artículos y determinantes en un 26 % (p Holm = .010, .017 y .049), sin efecto sobre las preposiciones ni sobre el orden de palabras (Tabla 17-7).

Tabla 17-7

Cambio en la tasa de errores por categoría: razones de tasas de incidencia relativas al control

Categoría	Contraste	Tasa pretest /100 palabras	RTI postest [IC 95 %]	p	p (Holm)	RTI diferido [IC 95 %]	p
Tiempo y aspecto verbal	RIA – C	2.07	0.66 [0.52, 0.85]	.001	.010	0.85 [0.66, 1.08]	.190
	RD – C	2.07	0.76 [0.60, 0.96]	.019	.115	0.77 [0.61, 0.97]	.025
Concordancia sujeto-verbo	RIA – C	1.53	0.70 [0.56, 0.88]	.002	.017	0.58 [0.45, 0.75]	< .001
	RD – C	1.53	0.82 [0.66, 1.03]	.083	.416	0.74 [0.56, 0.97]	.031
Artículos y determinantes	RIA – C	2.23	0.74 [0.60, 0.92]	.007	.049	0.77 [0.62, 0.95]	.016
	RD – C	2.23	0.89 [0.71, 1.11]	.288	.864	0.98 [0.78, 1.23]	.882
Preposiciones	RIA – C	1.71	0.93 [0.74, 1.17]	.541	.905	0.97 [0.77, 1.23]	.822
	RD – C	1.71	0.92 [0.73, 1.15]	.452	.905	0.85 [0.67, 1.08]	.189
Orden de palabras y estructura oracional	RIA – C	1.38	0.82 [0.64, 1.04]	.102	.416	0.99 [0.77, 1.27]	.925
	RD – C	1.38	0.67 [0.52, 0.87]	.002	.019	0.87 [0.66, 1.14]	.320

Nota: Ecuaciones de estimación generalizadas con distribución binomial negativa, enlace logarítmico, desplazamiento log(palabras/100) y correlación intercambiable dentro de estudiante. RTI = razón entre condiciones de la razón de tasas medición/pretest; valores < 1 indican una reducción de errores mayor que en el control. p (Holm) ajusta los diez contrastes del postest.

En el postest diferido se mantuvieron las reducciones de concordancia (RTI = 0.58 [0.45, 0.75], p < .001) y de artículos (RTI = 0.77 [0.62, 0.95], p = .016). La RD redujo sobre todo los errores de orden de palabras (RTI = 0.67 [0.52, 0.87], p Holm = .019) y, sin superar el ajuste, los de tiempo verbal (p = .019). La Figura 17-2 muestra las razones de tasas y las tasas observadas.

17.4.8 Mecanismo: el compromiso con la retroalimentación

El modelo de ecuaciones estructurales ajustó bien ($\chi^2(38) = 57.58$, p = .022; CFI = .990; TLI = .984; RMSEA = .035, IC 90 % [.014, .053]; SRMR = .037), con cargas estandarizadas de .75 en las tres subescalas de compromiso (Tabla 17-8).

Tabla 17-8

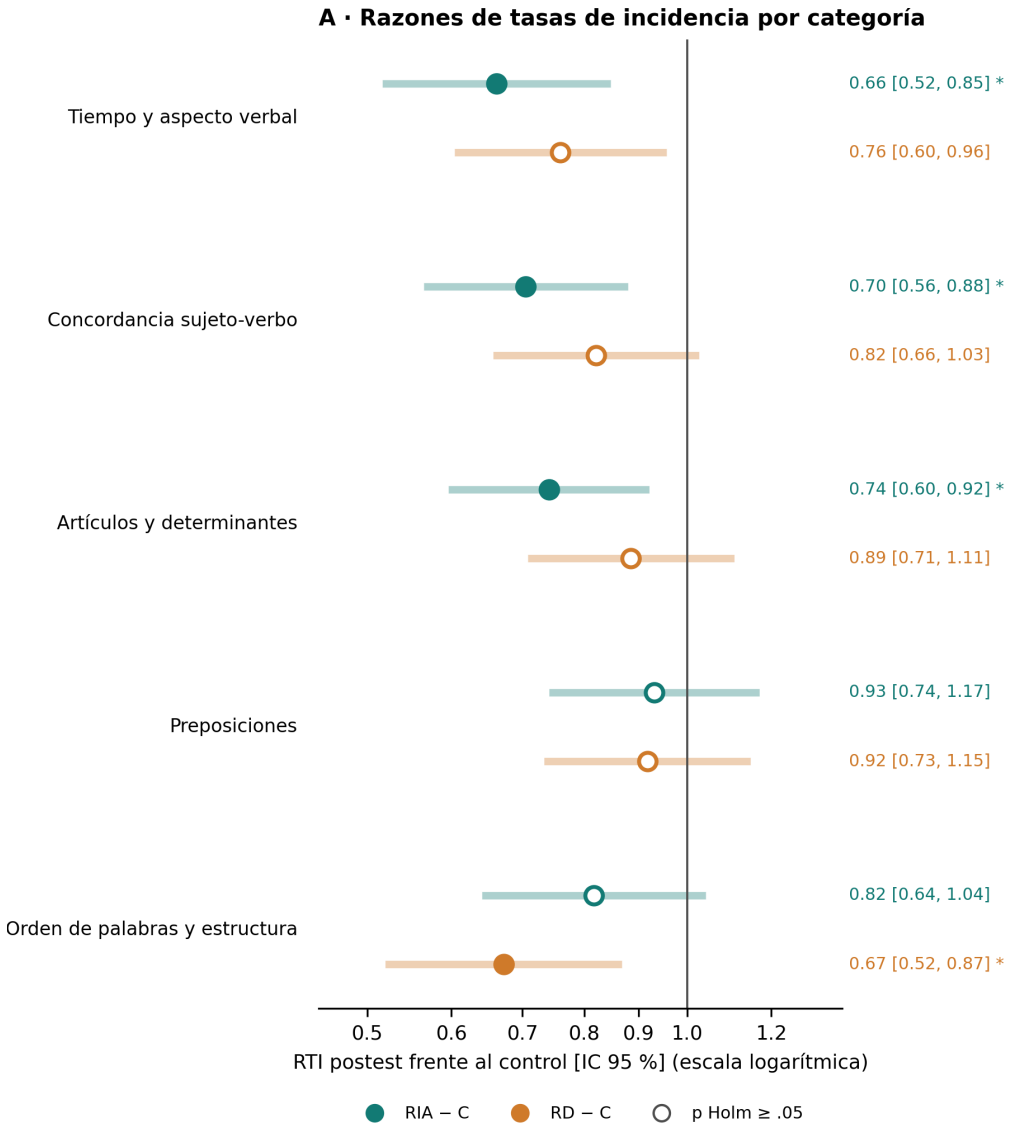
Modelo de ecuaciones estructurales del mecanismo: condición → compromiso con la retroalimentación → precisión sintáctica posterior (n = 419)

Parámetro	B (EE)	β	IC 95 %
RIA → compromiso (a_1)	0.44 (0.07)	.37	[0.30, 0.59]
RD → compromiso (a_2)	0.13 (0.07)	.11	[−0.01, 0.27]
Precisión previa → compromiso	0.08 (0.02)	.24	[0.04, 0.12]
Compromiso → precisión posterior (b)	0.72 (0.15)	.24	[0.43, 1.01]
RIA → precisión posterior, directo (c'_1)	0.70 (0.17)	.20	[0.38, 1.03]
RD → precisión posterior, directo (c'_2)	0.56 (0.16)	.16	[0.25, 0.87]
Precisión previa → precisión posterior	0.62 (0.05)	.64	[0.52, 0.71]
Efecto indirecto RIA → compromiso → precisión (a·b)	0.318	—	[0.169, 0.499] MC
Efecto total RIA (a·b + c′)	1.02	—	[0.66, 1.40] MC
Efecto indirecto RD → compromiso → precisión (a·b)	0.095	—	[−0.005, 0.214] MC
Efecto total RD (a·b + c′)	0.65	—	[0.33, 0.98] MC
R^2 compromiso / precisión posterior	.17 / .60	—	—

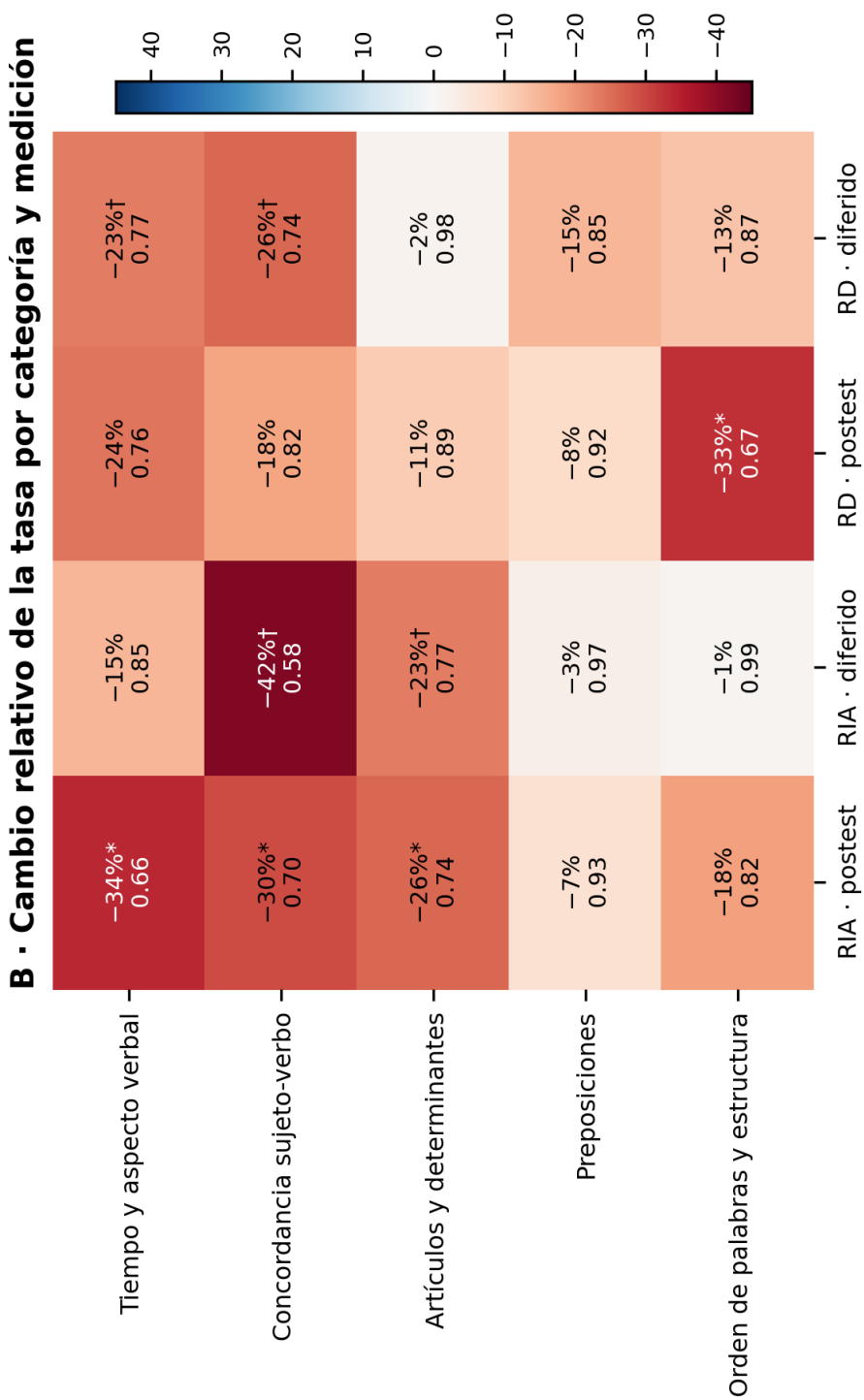
Nota: Estimación por máxima verosimilitud; los factores de precisión (pretest y postest) tienen tres indicadores (CSE/C y UTSE/UT ×10, E/100P invertido) y el de compromiso tres subescalas. MC = intervalo Monte Carlo con 20 000 extracciones (Preacher y Selig, 2012). Ajuste: $\chi^2(38) = 57.58$, p = .022; CFI = .990; TLI = .984; RMSEA = .035 [.014, .053]; SRMR = .037. B en la métrica del factor de precisión posterior (DE = 1.66).

Figura 17-2

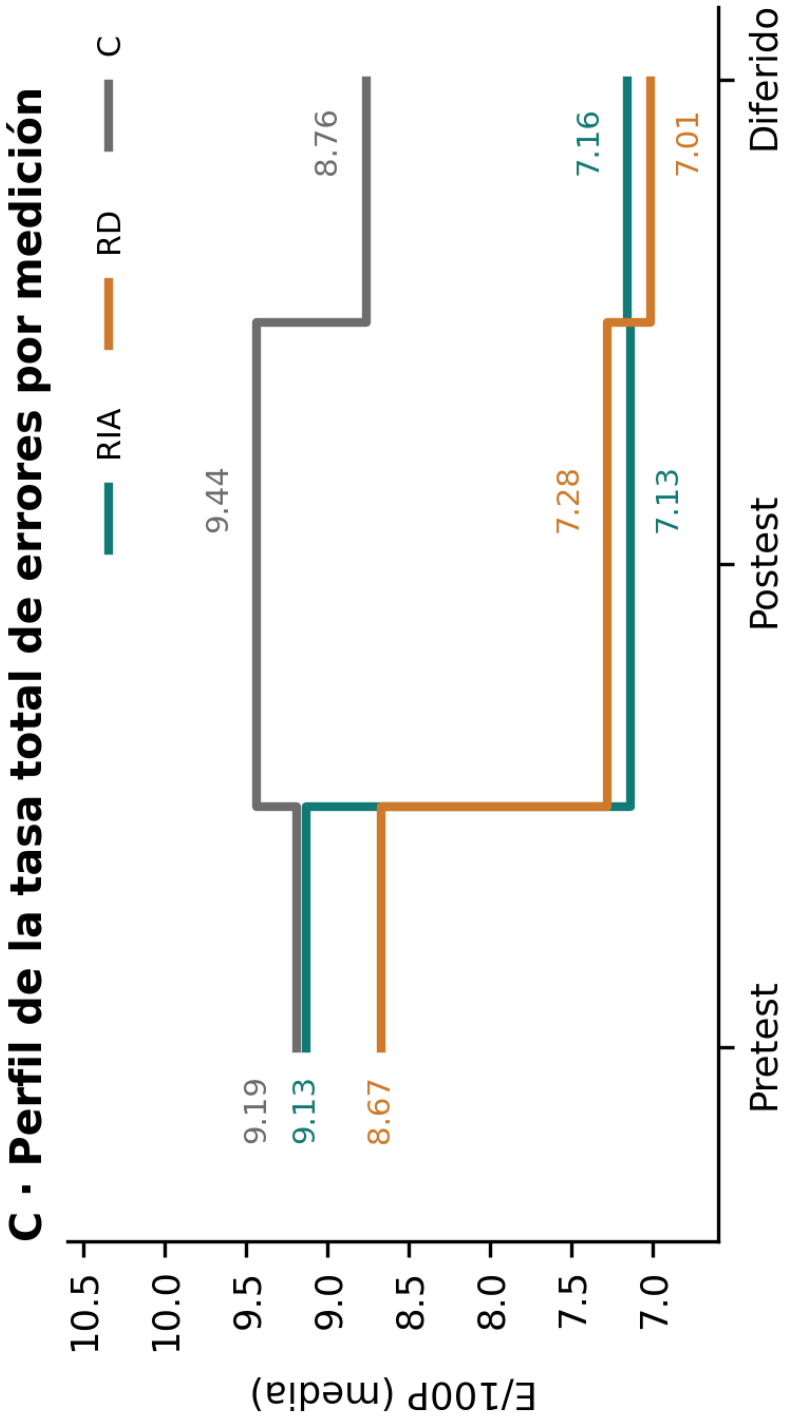
Tasas de error por categoría gramatical



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control; RTI = razón de tasas de incidencia. Estimaciones de ecuaciones de estimación generalizadas con distribución binomial negativa, desplazamiento log(palabras/100) y correlación intercambiable dentro de estudiante (Tabla 17-7). Panel A: RTI del postest de cada condición frente al control, con IC 95 % en escala logarítmica; valores menores que 1 indican menos errores que el control; relleno y asterisco = p Holm < .05 (RIA: tiempo y aspecto verbal 0.66 [0.52, 0.85], concordancia sujeto-verbo 0.70 [0.56, 0.88], artículos y determinantes 0.74 [0.60, 0.92]; RD: orden de palabras y estructura 0.67 [0.52, 0.87]).



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control; RTI = razón de tasas de incidencia. Estimaciones de ecuaciones de estimación generalizadas con distribución binomial negativa, desplazamiento log(palabras/100) y correlación intercambiable dentro de estudiante (Tabla 17-7). Panel C: media observada de errores por cada 100 palabras en cada medición (n por condición: 144 en el pretest; 141, 137 y 141 en el postest; 130, 129 y 130 en el diferido para RIA, RD y C).



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control; RTI = razón de tasas de incidencia. Estimaciones de ecuaciones de estimación generalizadas con distribución binomial negativa, desplazamiento log(palabras/100) y correlación intercambiable dentro de estudiante (Tabla 17-7). Panel C: media observada de errores por cada 100 palabras en cada medición (n por condición: 144 en el pretest; 141, 137 y 141 en el postest; 130, 129 y 130 en el diferido para RIA, RD y C).

La RIA aumentó el compromiso respecto del control (β = .37, p < .001; 0.79 DE del factor), mientras que el efecto de la RD fue menor y no significativo (β = .11, p = .061); el compromiso predijo la precisión posterior controlando la previa (β = .24, p < .001). El efecto indirecto de la RIA fue 0.32 (IC Monte Carlo 95 % [0.169, 0.499]; 0.19 DE del factor de precisión posterior), el 31 % de su efecto total, mientras que el de la RD incluyó el cero. Los efectos directos de ambas condiciones siguieron siendo significativos, lo que indica una mediación parcial del efecto de la RIA y sostiene H4. La varianza explicada de la precisión posterior fue R^2 = .60. La Figura 17-3 presenta la medida de precisión y el modelo del mecanismo.

17.4.9 Moderación por la competencia inicial

Tabla 17-9

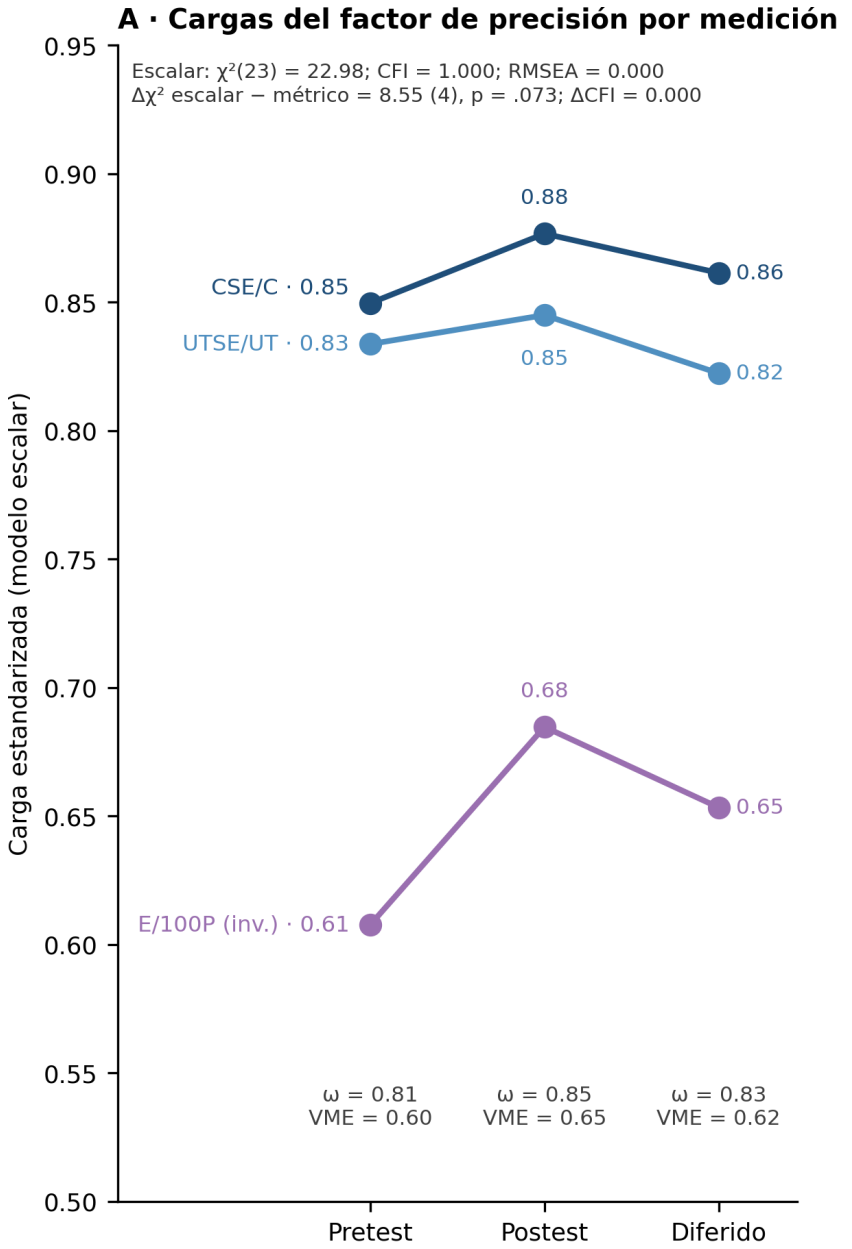
Moderación del efecto en el postest (CSE/C) por la competencia inicial en inglés

Contraste	Competencia inicial (puntaje)	Δ [IC 95 %] o región		g	p
RIA – C|	−1 DE (29.4)	0.134 [0.084, 0.185]		0.69	< .001
	Media (34.9)	0.097 [0.062, 0.133]		0.50	< .001
	+1 DE (40.3)	0.061 [0.009, 0.112]		0.31	.022
RD – C	−1 DE (29.4)	0.044 [−0.008, 0.096]		0.23	.095
	Media (34.9)	0.073 [0.037, 0.109]		0.38	< .001
	+1 DE (40.3)	0.101 [0.050, 0.152]		0.52	< .001
RIA × competencia	Interacción	−0.0368 [−0.0733, −0.0002]		—	.049
RD × competencia		0.0286 [−0.0081, 0.0652]		—	.127
RIA – C	Región de Johnson-Neyman	Efecto significativo por	debajo de 41.1 puntos	—	—
RD – C			encima de 30.2 puntos	—	—

Nota. Modelo lineal mixto del postest con intercepto aleatorio de sección, el pretest de CSE/C como covariable y la interacción condición × competencia inicial (prueba de ubicación, M = 34.9, DE = 5.4; rango observado 22-50). Región de Johnson-Neyman = valores de la competencia en los que el IC 95 % del efecto excluye el cero.

Figura 17-3

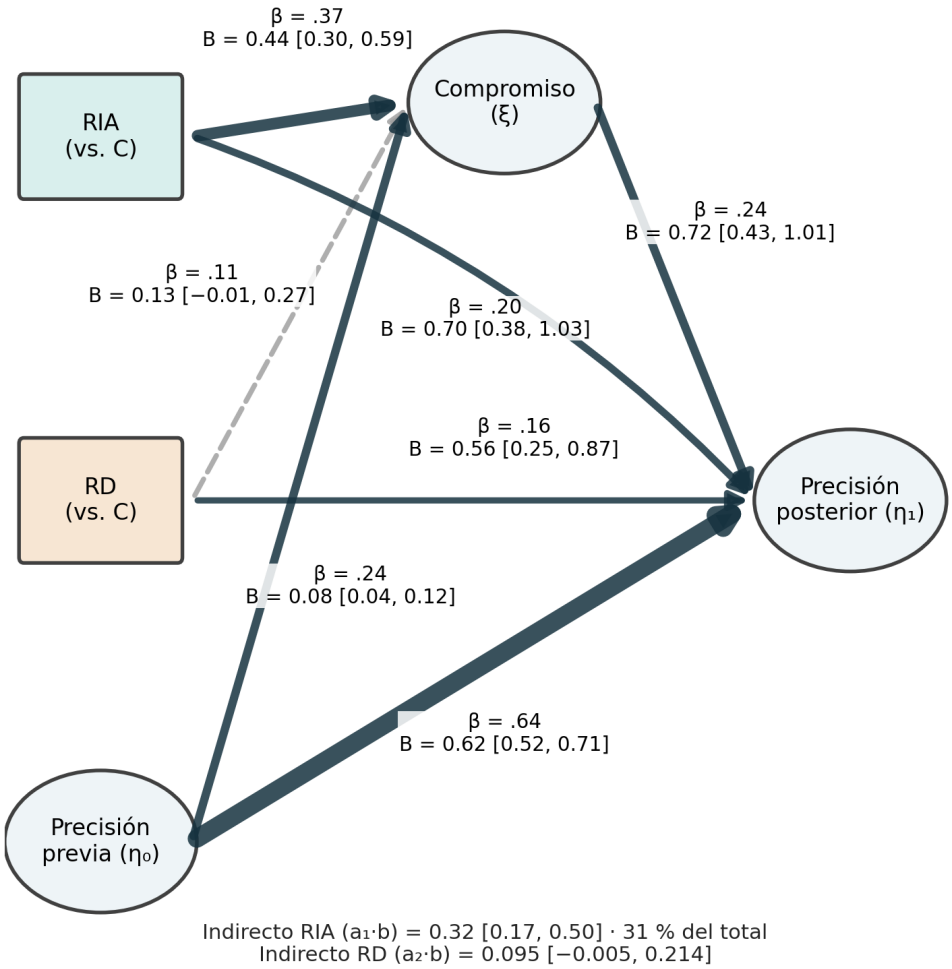
Medida de precisión y mecanismo de compromiso con la retroalimentación



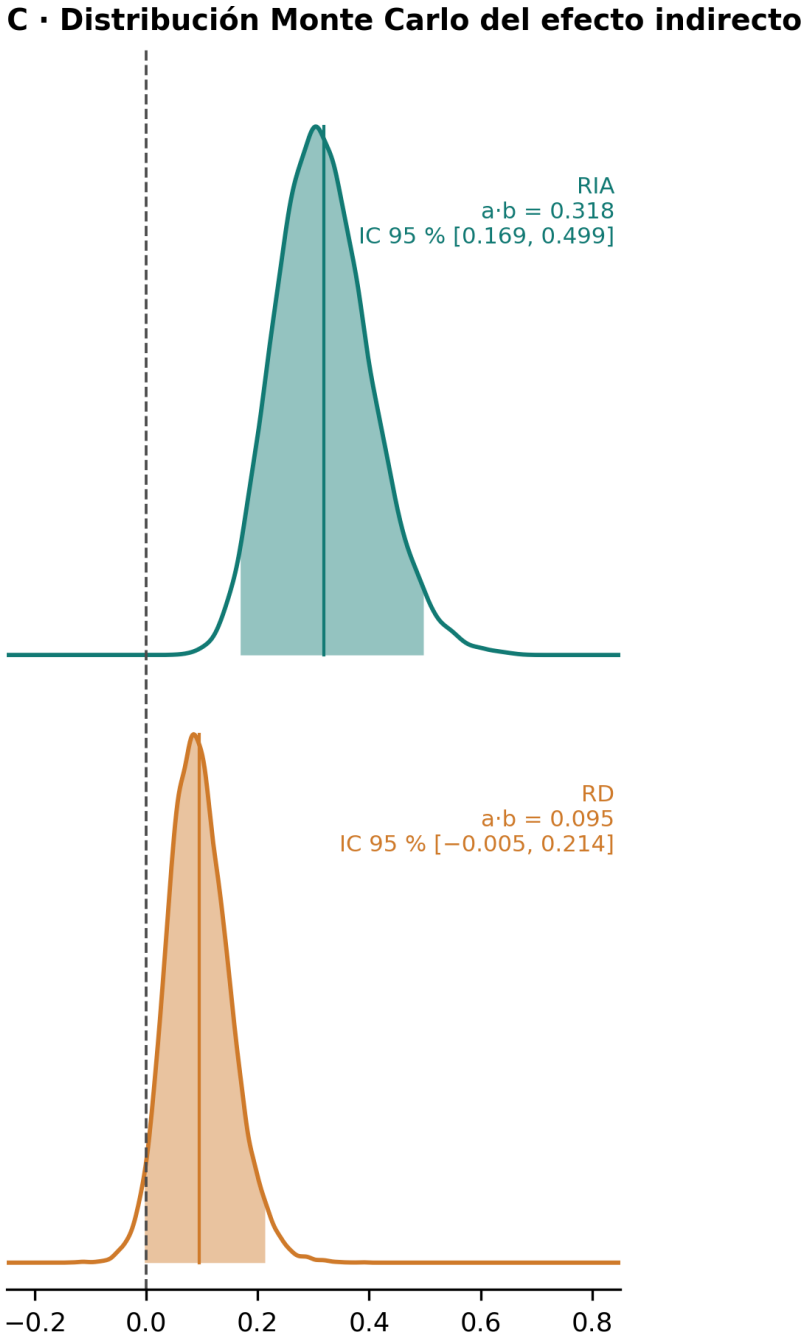
Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control. Panel A: cargas estandarizadas de CSE/C, UTSE/UT y E/100P (signo invertido) sobre el factor de precisión en cada medición, en el modelo de invarianza escalar estimado por máxima verosimilitud con los 389 estudiantes que completaron las tres mediciones; $\chi^2(23) = 22.98$, CFI = 1.000, RMSEA = 0.000; frente al modelo métrico, $\Delta\chi^2(4) = 8.55$, p = .073, ΔCFI = 0.000; ω = 0.81, 0.85 y 0.83 y varianza media extraída (VME) = 0.60, 0.65 y 0.62 en pretest, postest y diferido.

B · Modelo de ecuaciones estructurales del mecanismo

$\chi^2(38) = 57.58$ · CFI = 0.990 · TLI = 0.984 · RMSEA = 0.035 [0.014, 0.053] · SRMR = 0.037 · n =



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control. Panel B: modelo de ecuaciones estructurales del mecanismo (n = 419); coeficientes no estandarizados B con IC 95 % y estandarizados β; el grosor de la flecha es proporcional a |β| y las flechas discontinuas grises indican p ≥ .05 (RD → compromiso: B = 0.13 [−0.01, 0.27]). Ajuste: $\chi^2(38) = 57.58$, CFI = 0.990, TLI = 0.984, RMSEA = 0.035 [0.014, 0.053], SRMR = 0.037.



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control. Panel C: distribución de 20 000 extracciones Monte Carlo del efecto indirecto a·b; el área sombreada es el IC 95 % y la línea vertical la media. Indirecto de RIA = 0.318 [0.169, 0.499], 31 % del efecto total; indirecto de RD = 0.095 [−0.005, 0.214].

La interacción entre la RIA y la competencia inicial fue negativa y significativa (–0.037, EE = 0.019, p = .049): el efecto de la RIA sobre CSE/C fue mayor en los estudiantes con menor competencia y siguió siendo significativo por debajo de 41.1 puntos, un intervalo que incluye al 89 % de los participantes (Tabla 17-9). H5 se sostuvo. La RD mostró la tendencia opuesta, sin alcanzar significación (0.029, EE = 0.019, p = .127), y su efecto fue significativo solo por encima de 30.2 puntos.

17.4.10 Robustez

El efecto primario se mantuvo en todas las especificaciones alternativas (Tabla 17-10). Al excluir cada universidad, la estimación osciló entre 0.093 y 0.130, con intervalos que en todos los casos excluyeron el cero. La Figura 17-4 recoge la moderación y la robustez.

Tabla 17-10

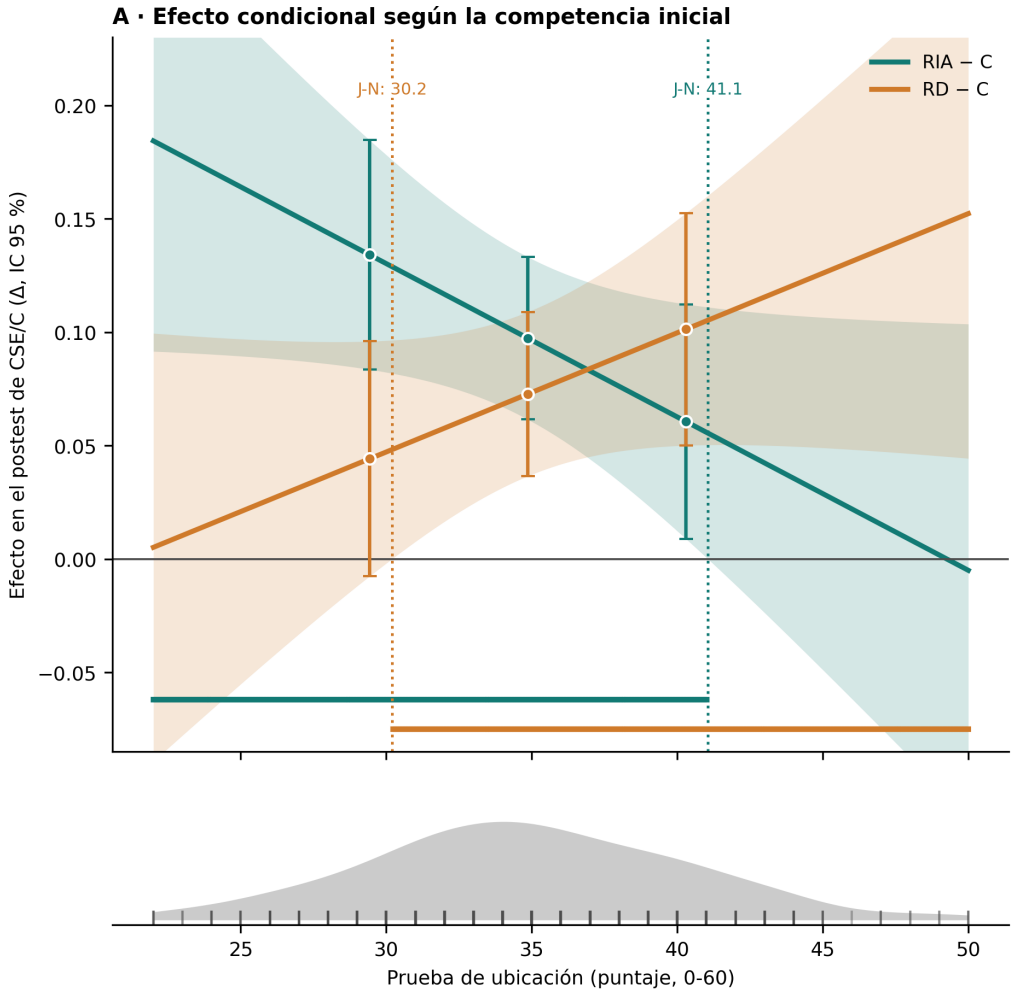
Robustez del efecto primario (RIA – C, postest, CSE/C) ante estimadores e inferencias alternativos

Especificación	Estimación [IC 95 %]	p
Modelo lineal mixto principal (estudiante dentro de sección, covariable de competencia)	Δ = 0.108 [0.064, 0.152]	
Modelo lineal mixto sin covariable de competencia	Δ = 0.108 [0.064, 0.152]	
ANCOVA del postest con errores robustos por sección (CR1, t con J – 1 = 23 gl)	Δ = 0.099 [0.056, 0.142]	
ANCOVA por rangos de Quade (RIA frente a control)	F(1, 280) = 29.62; A = .69	
Ecuaciones de estimación generalizadas binomiales (cláusulas sin error; razón de odds)	RO = 1.51 [1.26, 1.81]	< .001
Modelo lineal mixto con casos completos (n = 389)	Δ = 0.104 [0.058, 0.150]	
Modelo lineal mixto con pendiente aleatoria del tiempo por sección	Δ = 0.108 [0.064, 0.152]	
Sin la universidad de Lima	Δ = 0.093 [0.038, 0.147]	
Sin la universidad de Arequipa	Δ = 0.102 [0.048, 0.156]	
Sin la universidad de Trujillo	Δ = 0.130 [0.077, 0.182]	

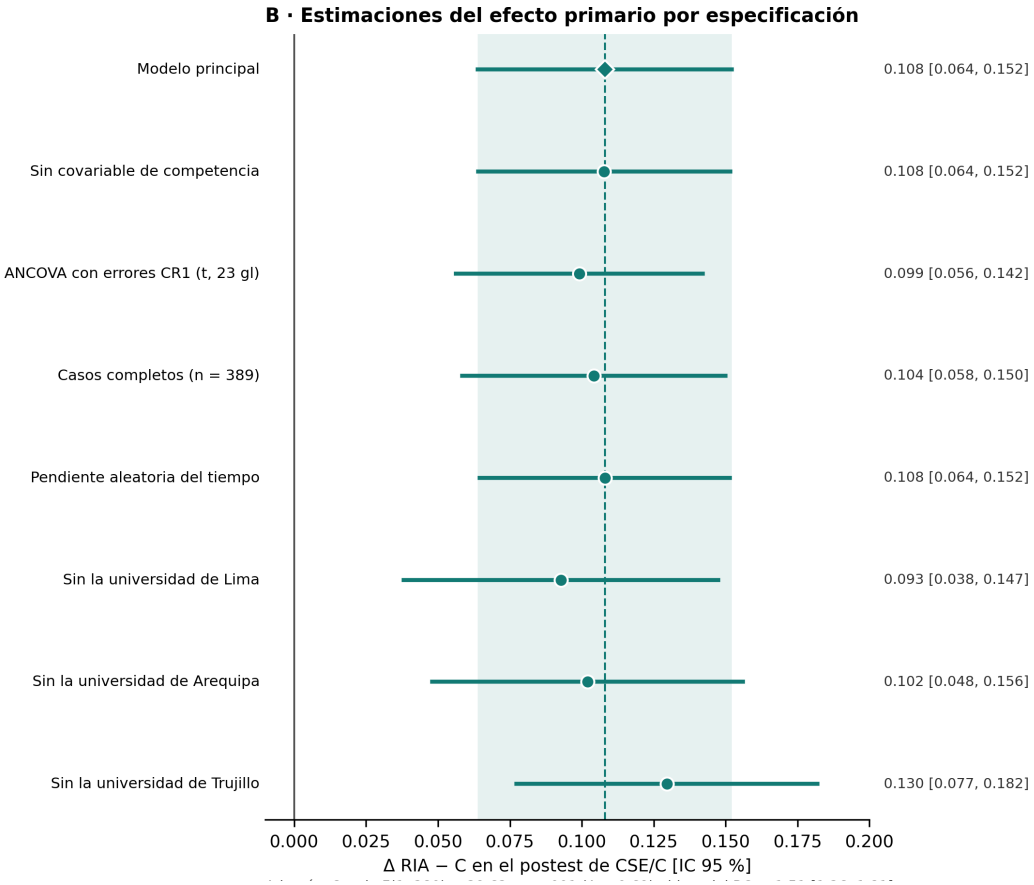
Nota: Δ = diferencia en la proporción de cláusulas sin error; RO = razón de odds de cláusula sin error; A = probabilidad de superioridad de los residuos de rango (Vargha-Delaney). CR1 = errores estándar robustos por conglomerado con la distribución t de J – 1 grados de libertad (J = 24 secciones).

Figura 17-4

Moderación por la competencia inicial y robustez del efecto primario



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control; CSE/C = proporción de cláusulas sin error. Panel A: efecto condicional de cada condición frente al control sobre CSE/C en el postest según el puntaje de la prueba de ubicación (modelo lineal mixto con interacción condición × competencia, ajustado por CSE/C del pretest; n = 419), con banda de IC 95 %, estimaciones puntuales a −1, 0 y +1 desviación típica (29.4, 34.9 y 40.3 puntos) y límites de Johnson-Neyman (líneas punteadas); las franjas inferiores marcan la región en la que el IC 95 % excluye el cero. Interacción RIA × competencia = −0.037, p = .049 (efecto significativo por debajo de 41.1 puntos, 89 % de la muestra); interacción RD × competencia = 0.029, p = .127 (efecto significativo por encima de 30.2 puntos). El panel inferior muestra la distribución de los puntajes de ubicación (N = 432).



Nota. RIA = retroalimentación con inteligencia artificial; RD = retroalimentación docente; C = control; CSE/C = proporción de cláusulas sin error. Panel B: diferencia RIA − C en el postest de CSE/C con IC 95 % en ocho especificaciones (Tabla 17-10); la línea discontinua y la banda sombreada corresponden al modelo principal, 0.108 [0.064, 0.152]; el intervalo más amplio es el de la exclusión de la universidad de Lima, 0.093 [0.038, 0.147]. La ANCOVA por rangos de Quade, $F(1, 280) = 29.62$, $p < .001$, y las ecuaciones de estimación generalizadas binomiales, razón de odds = 1.51 [1.26, 1.81], conducen a la misma conclusión.

17.5 DISCUSIÓN

17.5.1 Una persistencia que no sigue a la calidad de cada corrección

El hallazgo más desconcertante no es que ambas fuentes de retroalimentación superaran al control, sino la asimetría de su persistencia. La retroalimentación docente obtuvo una proporción mayor de revisiones adecuadas que la generativa (80 % frente a 72 %) y, sin embargo, perdió su ventaja en CSE/C a las ocho semanas, mientras que la RIA conservó la suya sobre el control en los tres índices, con una magnitud algo menor que en el postest. Si la durabilidad dependiera únicamente de la calidad de cada corrección, el patrón esperable sería el inverso. Los datos sugieren, en cambio, que lo decisivo es el procesamiento sostenido que la tarea obliga a realizar y que la RIA lo provocó con mayor intensidad, una lectura compatible con la evidencia longitudinal sobre cómo los estudiantes se implican con la retroalimentación automatizada (Xiaosa y Ping, 2025). Conviene matizarla: la diferencia entre ambas fuentes en el diferido no superó el ajuste por comparaciones múltiples, y que un contraste sea significativo y otro no, no basta para afirmar que difieran entre sí. Aun con esa cautela, el resultado ofrece una respuesta parcial a la objeción de que la mejora de un texto corregido no equivale a aprendizaje (Truscott, 2023): los textos del diferido se escribieron a mano, sin ayudas y sin acceso al asistente, y los estudiantes de la RIA seguían produciendo textos más precisos que los del control.

17.5.2 Magnitud y comparación entre fuentes

Respecto de la magnitud, el panorama es menos uniforme de lo que sugiere la literatura. Los efectos observados en CSE/C ($g = 0.56$) y en E/100P ($g = -0.57$) se sitúan en el rango de las síntesis sobre retroalimentación automatizada (Fleckenstein et al., 2023; Ngo et al., 2024), pero estas advierten que el beneficio sobre la gramática es menor que sobre el vocabulario, y aquí la precisión gramatical mejoró de forma apreciable. Una explicación tentativa es que una instrucción que exige una explicación metalingüística

breve y prohíbe la reescritura convierte al asistente en algo distinto de un corrector automático; esta conjetura no se contrastó directamente y debe tratarse como hipótesis. La cautela es pertinente porque otros trabajos centrados en la precisión con retroalimentación generativa hallaron mejoras parciales con errores persistentes (Pretorius y Thewissen, 2026; Fujisawa y Shintani, 2025), y porque los diseños previos más próximos al presente no permiten la misma inferencia. Wang et al. (2026) y Li y Collins (2026) contrastaron fuentes sin condición control, de modo que no pudieron separar la eficacia de cada una de la mera práctica de escritura; Asadi et al. (2025) y Khojasteh et al. (2025) estudiaron la combinación de la retroalimentación generativa con la docente y responden, por tanto, a una pregunta distinta de la de sustitución. Li y Collins informaron de que la calidad de la escritura mejoró entre borradores con las tres fuentes, lo que es consistente con la eficacia de ambas fuentes aquí observada, aunque la ausencia de control impide atribuir esa mejora a la retroalimentación. El diseño de tres brazos permite sostener que tanto la RIA como la RD superaron al control en el índice primario y que la RIA tendió a superar a la RD, con una ventaja que resistió el ajuste solo en UTSE/UT. La expresión «al menos igual» describe, por consiguiente, la ausencia de inferioridad observada y no una equivalencia probada, puesto que no se aplicó una prueba de equivalencia. Tampoco debe pasarse por alto que la ventaja de la RD sobre el control en CSE/C se sostiene en la familia confirmatoria, pero no cuando se corrigen los seis contrastes de ese índice.

En la dirección opuesta, los resultados concuerdan con la síntesis bayesiana más reciente y con el metaanálisis previo en que la retroalimentación correctiva escrita produce efectos moderados sobre la precisión (Kang y Han, 2015; Brown et al., 2026), y la magnitud de la RD, menor que la de la RIA pero significativa, encaja en ese marco.

17.5.3 El compromiso como mecanismo

El modelo estructural indica que alrededor de un tercio del efecto de la RIA se transmitió a través del compromiso con la retroalimentación, tras controlar la precisión inicial, mientras que la RD apenas modificó

el compromiso. El resultado extiende a la precisión sintáctica la relación entre retroalimentación generativa, compromiso y desempeño descrita con medidas globales (Teng y Huang, 2026) y respalda el enfoque que sitúa el compromiso en el centro del proceso (Zhang y Hyland, 2022; Ranalli, 2021). La mediación es parcial: la RIA mantuvo un efecto directo que podría deberse a la frecuencia y la inmediatez de las correcciones, aunque el estudio no permite aislar esa vía. Desde una perspectiva aplicada, la implicación es que el valor de la retroalimentación generativa residiría menos en la herramienta que en una tarea que obliga a procesar la explicación y a revisar. En esa línea, el procesamiento individual o colaborativo de la retroalimentación influye en la precisión y en la revisión (Mujtaba et al., 2021; Peng, 2024), y una práctica apoyada en retroalimentación generativa puede incidir en la alfabetización en retroalimentación de los estudiantes (Lu et al., 2026). Las salvaguardas contra la reescritura completa, como la entrega del registro de conversación, serían así parte de la intervención y no un detalle de implementación (Pretorius y Thewissen, 2026).

17.5.4 Qué errores cambian y para quién

La reducción de errores de la RIA se concentró en tiempo verbal, concordancia y artículos, categorías gobernadas por reglas, y no alcanzó a las preposiciones ni al orden de palabras, un patrón consistente con la mayor sensibilidad de los errores tratables a la retroalimentación (Mujtaba et al., 2021; Peng, 2024). La RD, en cambio, redujo los errores de orden de palabras; cabe conjeturar que una explicación docente aborda mejor lo que depende de la estructura de la oración que una explicación breve por error. Puesto que el análisis por categoría fue exploratorio, la complementariedad que esto insinúa debe confirmarse con muestras independientes. Por lo que respecta a la competencia inicial, el efecto de la RIA fue mayor en quienes partían con menos dominio, mientras que el de la RD tendió a concentrarse en los más competentes; sin embargo, la interacción de la RIA apenas alcanzó la significación y la de la RD no la alcanzó, de modo que la evidencia es sugestiva y no concluyente. De confirmarse, una fuente que beneficia más a quienes parten con menor

dominio tendría un valor de equidad particular en grupos de competencia heterogénea.

17.5.5 Implicaciones para el contexto peruano

En el contexto peruano, donde el inglés se enseña con pocas horas y grupos numerosos (Saez et al., 2025; La Torre, 2021), los resultados sugieren que la retroalimentación generativa puede ampliar el alcance de la retroalimentación sobre la forma sin pérdida de eficacia frente a la docente, siempre que se integre con instrucciones que fuercen el procesamiento y con alfabetización crítica sobre su uso (Darvin, 2025; Wang y Wang, 2025). Esta condición no es accesoria: la literatura sobre alfabetización digital sugiere que el compromiso con estas herramientas es heterogéneo y depende de lo que el estudiante sabe hacer con ellas (Zhang y Hyland, 2025), y el mismo asistente puede producir aprendizajes muy distintos según cómo se incorpore a la tarea.

17.5.6 Limitaciones y líneas futuras

El estudio tiene limitaciones. La selección de universidades y secciones fue no probabilística, por lo que la generalización descansa en la replicación. La asignación individual dentro de las secciones expone a contaminación entre condiciones si los estudiantes comparten la retroalimentación; las instrucciones y el registro del uso la limitaron, pero no la eliminan, y su efecto esperado es atenuar las diferencias. El compromiso se midió una sola vez, al final de la intervención, de modo que el modelo del mecanismo no establece la precedencia temporal entre compromiso y cambio en la precisión; los estudios futuros deberían medirlo durante la intervención. La ventaja de la RIA sobre la RD no fue consistente entre índices tras el ajuste por comparaciones múltiples, no se realizó una prueba formal de equivalencia y los efectos por categoría de error deben confirmarse. Por último, el asistente se configuró con una instrucción fija y un modelo concreto; la rápida evolución de estas herramientas exige replicar el diseño con otras configuraciones (Liu et al., 2025; Yang y Chen, 2026).

17.6 CONCLUSIONES

Este ensayo aleatorizado de tres brazos aporta evidencia de que la retroalimentación con inteligencia artificial generativa, configurada para señalar errores, explicarlos y no reescribir el texto, mejora la precisión sintáctica de la escritura en inglés de estudiantes universitarios peruanos más que la práctica sin retroalimentación sobre la forma. Las hipótesis H1 y H2 se sostuvieron: ambas fuentes superaron al control en el postest en CSE/C, y la generativa mostró una magnitud al menos comparable a la docente, sin que su ventaja sobre ella resistiera de forma consistente el ajuste por comparaciones múltiples. La hipótesis H3 también se sostuvo, ya que la ventaja de la RIA persistió ocho semanas después, cuando la de la RD dejó de ser significativa en la medida principal.

Alrededor de un tercio del efecto de la RIA se transmitió a través del compromiso del estudiante con la retroalimentación, de modo que H4 se confirmó con una mediación parcial, y su beneficio fue mayor para quienes partían con menor competencia, como planteaba H5. La reducción de errores se concentró en tiempos verbales, concordancia y artículos, y la retroalimentación docente conservó un papel propio en el orden de palabras.

Para las instituciones, la implicación práctica es que la retroalimentación generativa puede complementar a la docente y ampliar su alcance, a condición de que se diseñe para provocar compromiso y se acompañe de salvaguardas contra la reescritura completa. Para la investigación, quedan abiertas la medición repetida del compromiso durante la intervención, la confirmación independiente de los efectos por categoría, la comprobación formal de equivalencia con la retroalimentación docente y la replicación con otras configuraciones del asistente y con poblaciones distintas, dado que la selección de secciones fue no probabilística.

17.7 REFERENCIAS BIBLIOGRÁFICAS

Abudalfa, S., & Barrot, J. (2026). Generative artificial intelligence for automated writing evaluation: A systematic review of trends, efficacy, and challenges. *Assessing Writing, 68,* 101041. https://doi.org/10.1016/j.asw.2026.101041

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity, 56,* 101766. https://doi.org/10.1016/j.tsc.2025.101766

Banihashem, S., Kerman, N., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1). https://doi.org/10.1186/s41239-024-00455-4

Barrot, J. (2023). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(4), 584-607. https://doi.org/10.1080/09588221.2021.1936071

Barrot, J., & Agdeppa, J. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing, 47,* 100510. https://doi.org/10.1016/j.asw.2020.100510

Brown, D., Liu, Q., & Norouzian, R. (2026). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research, 30*(3), 1357-1389. https://doi.org/10.1177/13621688221147374

Chen, F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing*, *67*, 101186. https://doi.org/10.1016/j.jslw.2025.101186

Fan, N. (2023). Exploring the Effects of Automated Written Corrective Feedback on EFL Students' Writing Quality: A Mixed-Methods Study. *Sage Open*, *13*(2). https://doi.org/10.1177/21582440231181296

Fleckenstein, J., Liebenow, L., & Meyer, J. (2023). Automated feedback and writing: a multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1162454

Fujisawa, Y., & Shintani, N. (2025). Comparative effects of direct written corrective feedback and AI-generated reformulation on second language learners' depth of processing and grammatical accuracy. *System*, *133*, 103765. https://doi.org/10.1016/j.system.2025.103765

Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, *35*(9), 2312-2331. https://doi.org/10.1080/09588221.2021.1879161

Gutiérrez, V. (2024). Impact of a YouTube Channel on ESL Students' Oral Skills in a Peruvian Higher Institute. *World Journal of English Language*, *14*(5), 392. https://doi.org/10.5430/wjel.v14n5p392

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, X., Xu, W., Li, F., & Yu, Z. (2024). A Meta-analysis of Effects of Automated Writing Evaluation on Anxiety, Motivation, and Second Language Writing Skills. *The Asia-Pacific Education Researcher, 33*(4), 957-976. https://doi.org/10.1007/s40299-024-00865-y

Hwang, H., Chang, X., & Sun, J. (2025). Generative AI is useful for second language writing, but when, why, and for how long do learners use it? *Journal of Second Language Writing, 69,* 101230. https://doi.org/10.1016/j.jslw.2025.101230

Kailin, Z., & Saeed, M. (2026). Chinese EFL learners' engagement with ChatGPT feedback on academic writing: A case study in Malaysia. *Computers and Composition, 79,* 102976. https://doi.org/10.1016/j.compcom.2025.102976

Kang, E., & Han, Z. (2015). The Efficacy of Written Corrective Feedback in Improving L2 Written Accuracy: A Meta-Analysis. *The Modern Language Journal, 99*(1), 1-18. https://doi.org/10.1111/modl.12189

Kang, J., & Yi, Y. (2023). Beyond ChatGPT: Multimodal generative AI for L2 writers. *Journal of Second Language Writing, 62,* 101070. https://doi.org/10.1016/j.jslw.2023.101070

Khojasteh, L., Soori, A., & Javed, F. (2025). Comparing teacher E-feedback, AI feedback, and hybrid feedback in enhancing EFL writing skills.

Technology in Language Teaching & Learning, 7(3), 102626. https://doi.org/10.29140/tltl.v7n3.102626

Kim, M., & Chon, Y. (2025). The impact of self-revision, machine translation, and ChatGPT on L2 writing: Raters' assessments, linguistic complexity, and error correction. *Assessing Writing, 65,* 100950. https://doi.org/10.1016/j.asw.2025.100950

Koo, T., & Li, M. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

La Torre, C. (2021). Virtual Classroom Usage and User Perception for English Learning as a Second Language at Universities in Lima, Peru. *International Journal of Emerging Technologies in Learning (iJET), 16*(08), 261. https://doi.org/10.3991/ijet.v16i08.19221

Li, A., & Collins, P. (2026). Formative feedback across sources: Student perceptions and writing outcomes with instructor, peer, and AI-generated feedback. *Reading and Writing, 39*(8), 3107-3129. https://doi.org/10.1007/s11145-026-10761-0

Liu, Y., Hau, K., & Liu, H. (2024). Linear Mixed-Effects Models for Dependent Data: Power and Accuracy in Parameter Estimation. *Multivariate Behavioral Research, 59*(5), 978-994. https://doi.org/10.1080/00273171.2024.2350236

Liu, Y., Lu, X., & Qi, H. (2025). Comparing GPT-based approaches in automated writing evaluation. *Assessing Writing, 66,* 100961. https://doi.org/10.1016/j.asw.2025.100961

Lu, Q., Yao, Y., Xiao, L., Zhu, X., & Yin, H. (2026). Exploring the impact of a GenAI-supported feedback practice on student feedback literacy in L2 writing. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2025.2605541

Luong, R., & Flake, J. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods, 28*(4), 905-924. https://doi.org/10.1037/met0000441

Ma, X., Zhang, C., & Lee, I. (2026). L2 learners' engagement with AI-generated feedback on writing. *Assessing Writing*, *68*, 101020. https://doi.org/10.1016/j.asw.2026.101020

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989x.1.2.130

Mackinnon, S., Curtis, R., & O'Connor, R. (2022). Tutorial in Longitudinal Measurement Invariance and Cross-lagged Panel Models Using Lavaan. *Meta-Psychology*, *6*. https://doi.org/10.15626/mp.2020.2595

Mujtaba, S., Reynolds, B., Parkash, R., & Singh, M. (2021). Individual and collaborative processing of written corrective feedback affects second language writing accuracy and revision. *Assessing Writing*, *50*, 100566. https://doi.org/10.1016/j.asw.2021.100566

Ngo, T., Chen, H., & Lai, K. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis.

Interactive Learning Environments, 32(2), 727-744. https://doi.org/10.1080/10494820.2022.2096642

Nicklin, C., & Vitta, J. (2021). Effect-Driven Sample Sizes in Second Language Instructed Vocabulary Acquisition Research. *The Modern Language Journal, 105*(1), 218-236. https://doi.org/10.1111/modl.12692

Peng, C. (2024). Beyond accuracy gains: Investigating the impact of individual and collaborative feedback processing on L2 writing development. *Assessing Writing, 61*, 100876. https://doi.org/10.1016/j.asw.2024.100876

Phuoc, V., & Barrot, J. (2022). Complexity, accuracy, and fluency in L2 writing across proficiency levels: A matter of L1 background? *Assessing Writing, 54*, 100673. https://doi.org/10.1016/j.asw.2022.100673

Plonsky, L., & Oswald, F. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186x.2024.2410101

Preacher, K., & Selig, J. (2012). Advantages of Monte Carlo Confidence Intervals for Indirect Effects. *Communication Methods and Measures, 6*(2), 77-98. https://doi.org/10.1080/19312458.2012.679848

Pretorius, M., & Thewissen, J. (2026). Mediating L2 writing revision: The impact of teacher-moderated ChatGPT feedback on L2 accuracy and lexical diversity. *Journal of Second Language Writing, 71*, 101286. https://doi.org/10.1016/j.jslw.2026.101286

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816

Saez, N., Cunza, D., Abanto, C., & Montalvo, D. (2025). Relationship between English self-efficacy and language learning strategies among Peruvian university students: the mediating role of academic self-efficacy. *Frontiers in Education, 10.* https://doi.org/10.3389/feduc.2025.1668300

Tan, S., Cho, Y., & Xu, W. (2023). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments, 31*(10), 7276-7286. https://doi.org/10.1080/10494820.2022.2066137

Teng, M. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M., & Huang, J. (2026). Assessing ChatGPT Feedback for EFL Learners' Engagement and Writing Performance. *International Journal of Applied Linguistics.* https://doi.org/10.1111/ijal.70155

Truscott, J. (2023). What about validity? Thoughts on the state of research on written corrective feedback. *Feedback Research in Second Language, 1,* 33-53. https://doi.org/10.32038/frsl.2023.01.03

Vitta, J., Nicklin, C., & McLean, S. (2022). Effect Size–Driven Sample-Size Planning, Randomization, and Multisite Use in L2 Instructed Vocabulary Acquisition Experimental Samples. *Studies in Second Language Acquisition, 44*(5), 1424-1448. https://doi.org/10.1017/s0272263121000541

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67,* 101187. https://doi.org/10.1016/j.jslw.2025.101187

Wang, Y., Hu, W., & Guo, Q. (2026). GPT versus teacher feedback on English writing: An analysis of feedback characteristics and student engagement. *System, 138,* 103967. https://doi.org/10.1016/j.system.2025.103967

Xiaosa, L., & Ping, K. (2025). How L2 student writers engage with automated feedback: A longitudinal perspective. *Assessing Writing, 64,* 100919. https://doi.org/10.1016/j.asw.2025.100919

Xu, Z., Gao, F., Fa, A., Qu, W., & Zhang, Z. (2024). Statistical power analysis and sample size planning for moderated mediation models. *Behavior Research Methods, 56*(6), 6130-6149. https://doi.org/10.3758/s13428-024-02342-2

Yang, C., & Chen, H. (2026). ChatGPT and L2 Chinese writing: evaluating the impact of model version and prompt language on automated

corrective feedback. *Computer Assisted Language Learning*, *39*(4), 1119-1147. https://doi.org/10.1080/09588221.2025.2453205

Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing*, *68*, 101203. https://doi.org/10.1016/j.jslw.2025.101203

Zhai, N., & Ma, X. (2023). The Effectiveness of Automated Writing Evaluation on Writing Quality: A Meta-Analysis. *Journal of Educational Computing Research*, *61*(4), 875-900. https://doi.org/10.1177/07356331221127300

Zhai, Y. (2025). Automated written corrective feedback: fostering writing accuracy, self-regulation, and motivational dynamics in EFL learners. *BMC Psychology*, *14*(1). https://doi.org/10.1186/s40359-025-03765-y

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zhang, Z., & Hyland, K. (2025). The role of digital literacy in student engagement with automated writing evaluation (AWE) feedback on second language writing. *Computer Assisted Language Learning*, *38*(5-6), 1060-1085. https://doi.org/10.1080/09588221.2023.2256815

CAPÍTULO XVIII

ENSEÑANZA DEL INGLÉS COMO TERCERA LENGUA EN ESCUELAS INTERCULTURALES BILINGÜES RURALES ANDINAS DEL PERÚ: PRÁCTICAS CULTURALMENTE PERTINENTES, BARRERAS CONTEXTUALES, AUTOEFICACIA Y DESEMPEÑO EN UN MODELO DE ECUACIONES ESTRUCTURALES MULTIGRUPO, 2025

Diego Alejo Morales Aguero

Universidad nacional santiago antúnez de mayolo

Resumen

Se estimó en qué medida las prácticas didácticas culturalmente pertinentes y las barreras contextuales de la escuela intercultural bilingüe rural andina se relacionan con el desempeño en inglés como tercera lengua, qué parte de esa relación se transmite por la autoeficacia y si el modelo es equivalente entre estudiantes quechuahablantes y aimarahablantes. Se realizó un estudio transversal con 480 estudiantes de 3.º a 5.º de secundaria de 32 instituciones rurales del sur andino del Perú (240 por lengua), evaluados en 2025 con escalas de prácticas, barreras y autoeficacia y con tres pruebas de desempeño. Se ajustó un modelo de ecuaciones estructurales por máxima verosimilitud, con bootstrap de conglomerados y una secuencia de invarianza multigrupo. El modelo ajustó a los datos (CFI = .991; RMSEA = .024; SRMR = .034). Las prácticas se asociaron con mayor autoeficacia (β = .335) y esta con el desempeño (β = .393); el efecto indirecto de las prácticas fue .132 (IC 95 % [.077, .195]), la mitad de su efecto total. Las barreras tuvieron un efecto total negativo (–.139; IC 95 % [–.258, –.034]), pero ni su trayectoria hacia la autoeficacia ni su efecto indirecto se distinguieron de cero. La medición y las trayectorias fueron invariantes entre ambas lenguas. La autoeficacia aparece como el mecanismo por el que la pertinencia cultural de la didáctica se relaciona con el aprendizaje de una tercera lengua; las barreras se asocian con menor desempeño por vías no distinguibles a través de las creencias de eficacia, y un mismo modelo describe a ambas comunidades.

Palabras clave: inglés como tercera lengua; educación intercultural bilingüe; autoeficacia; enseñanza culturalmente pertinente; invarianza de medición; ecuaciones estructurales.

Abstract

This study estimated how far culturally relevant teaching practices and contextual barriers in rural Andean intercultural bilingual schools relate to students' English performance as a third language, how much of that relation is transmitted

through English self-efficacy, and whether the model holds equally for Quechua- and Aymara-speaking students. A cross-sectional study was conducted with 480 students in grades 9 to 11 from 32 rural institutions in the southern Peruvian Andes (240 per language), assessed in 2025 with scales of practices, barriers and self-efficacy and three performance tests. A latent-variable structural equation model was fitted by maximum likelihood, with cluster bootstrap standard errors and a multigroup invariance sequence. The model fitted the data (CFI = .991; RMSEA = .024; SRMR = .034). Practices were associated with higher self-efficacy (β = .335), and self-efficacy with performance (β = .393); the indirect effect of practices was .132 (95% CI [.077, .195]), half of their total effect. Barriers had a negative total effect (−.139; 95% CI [−.258, −.034]), but neither their path to self-efficacy nor their indirect effect differed from zero. Measurement and structural paths were invariant across the two language groups. Self-efficacy emerges as the mechanism through which the cultural relevance of teaching relates to third-language learning; school barriers are associated with lower performance through routes not distinguishable via efficacy beliefs, and the same model describes both Indigenous-language communities.

Keywords: *English as a third language; intercultural bilingual education; self-efficacy; culturally relevant teaching; measurement invariance; structural equation modeling.*

18.1 INTRODUCCIÓN

18.1.1 Planteamiento del problema

La enseñanza del inglés se ha vuelto obligatoria en la educación secundaria pública de numerosos países de América Latina, incluidas sus zonas rurales e indígenas, dentro de una política de promoción del inglés extendida en toda la región (Fabre et al., 2026; Izquierdo et al., 2021). Con ello llega a escuelas cuyos estudiantes no aprenden el inglés como segunda lengua, sino como tercera: después de una lengua originaria aprendida en el hogar y del castellano escolar (Fabre et al., 2025). En el Perú, la educación intercultural bilingüe (EIB) se practica sobre todo en

contextos rurales (Flores et al., 2023) y atiende, entre otras, a comunidades quechuahablantes y aimarahablantes de la sierra sur, cuyos estudiantes cursan además el inglés como área obligatoria de la educación secundaria. La escuela EIB rural andina reúne así, en un mismo aula, tres lenguas con estatus sociales muy desiguales y una oferta didáctica pensada, en su origen, para estudiantes monolingües urbanos (Sullón, 2024).

Los estudios disponibles describen con detalle las dificultades de este escenario: docentes generalistas o sin formación en la especialidad, escasez de recursos y de conectividad, y una débil articulación de la enseñanza con la cultura local (Izquierdo et al., 2021; Ortiz y Contreras, 2023; Duque et al., 2024; Vega et al., 2025; Guevara et al., 2026). El propio sistema EIB peruano reconoce problemas estructurales de formación docente y de financiamiento que limitan la atención a la diversidad cultural (Moreano et al., 2026; Tarrillo et al., 2026). Sin embargo, esta evidencia es mayoritariamente cualitativa o descriptiva: documenta los obstáculos, pero no estima cuánto pesa cada uno en el aprendizaje del inglés ni a través de qué mecanismo psicológico actúa.

18.1.2 Antecedentes y vacío de conocimiento

La investigación en adquisición de segundas lenguas ofrece un candidato sólido para ese mecanismo: la autoeficacia, es decir, el juicio del estudiante sobre su capacidad para alcanzar metas de aprendizaje en la lengua meta. Un metaanálisis de 40 muestras independientes (N = 23 050) estimó una correlación media de r = .46 entre autoeficacia en la segunda lengua y logro, con variaciones sistemáticas según la lengua materna y la lengua meta de los aprendices (Goetze y Driver, 2022), y la relación se ha replicado para la escritura (Sun et al., 2021). En el Perú se dispone de una escala de autoeficacia en inglés con invarianza factorial en estudiantes universitarios (Mendoza et al., 2023) y de una escala de autoeficacia en la escritura validada en estudiantes quechuahablantes de educación básica (Curo et al., 2025). Las emociones de aula acompañan a ese mecanismo: el disfrute predice el dominio oral del inglés en futuros docentes peruanos (Guerra et al., 2024), y la ansiedad ante la lengua extranjera es medible en

la secundaria pública del sur del país (Quispe et al., 2024).

Tres limitaciones impiden trasladar estos hallazgos a la escuela EIB rural andina. Primera, casi toda la evidencia procede de estudiantes universitarios o urbanos para quienes el inglés es la segunda lengua; la enseñanza del inglés como tercera lengua a estudiantes indígenas plantea condiciones propias de identidad y de alfabetización previa en la lengua originaria y en el castellano (Fabre et al., 2025), y la escuela rural peruana presenta brechas de rendimiento frente a la urbana que se explican sobre todo por diferencias de recursos socioeconómicos y escolares (Rojas et al., 2024). Segunda, la enseñanza culturalmente pertinente se ha examinado en el inglés rural solo desde la perspectiva docente y con diseños cualitativos o descriptivos (Vega et al., 2025; Montoya et al., 2025), sin estimar su efecto sobre las creencias de eficacia de los estudiantes. Tercera, ningún estudio ha comprobado si las relaciones entre condiciones de enseñanza, autoeficacia y desempeño en inglés se sostienen por igual entre comunidades de lengua originaria distintas, condición necesaria para generalizar cualquier recomendación pedagógica dentro de un sistema EIB que atiende a varias lenguas (Kvietok y Hornberger, 2023), y más aún cuando el dominio del inglés varía de forma marcada entre tipos de escuela y regiones (Elizondo et al., 2026).

18.1.3 Pregunta, objetivos y aporte

Este estudio responde a la pregunta: ¿en qué medida las prácticas didácticas culturalmente pertinentes y las barreras contextuales de la escuela EIB rural andina se relacionan con el desempeño en inglés de sus estudiantes, y qué parte de esa relación se transmite a través de la autoeficacia en inglés? Los objetivos son

a) Estimar las trayectorias directas e indirectas de un modelo de ecuaciones estructurales con dos condiciones de enseñanza exógenas, un mediador y un desempeño medido con pruebas, y

b) Comprobar la invarianza de la medición y de las trayectorias estructurales entre estudiantes quechuahablantes y aimarahablantes.

El aporte es teórico y metodológico, y trasciende el caso peruano. En lo teórico, el estudio sitúa la autoeficacia —un mecanismo establecido para la segunda lengua— como mediador en el aprendizaje de una tercera lengua en contextos de lengua minorizada, y separa dos fuentes de variación que la literatura suele confundir: lo que el docente hace en el aula (prácticas pertinentes) y lo que el entorno escolar impide (barreras). En lo metodológico, ofrece una prueba explícita de invarianza entre dos comunidades lingüísticas originarias, un tipo de evidencia que la investigación sobre aprendizaje de lenguas en poblaciones indígenas casi nunca aporta y que es la condición para comparar grupos o agregar resultados entre ellos. La escuela rural andina del Perú funciona aquí como sitio de prueba de un modelo aplicable a cualquier sistema educativo que enseñe una lengua internacional a estudiantes cuya lengua materna es originaria.

18.2 MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

18.2.1 La autoeficacia como mecanismo del aprendizaje de lenguas adicionales

La teoría social cognitiva sostiene que las creencias de eficacia se construyen a partir de experiencias de dominio, observación de modelos, persuasión y estados emocionales, y que regulan el esfuerzo y la persistencia ante tareas difíciles. En el aprendizaje de lenguas, esta regulación es especialmente relevante porque el progreso es lento y la exposición fuera del aula es mínima; de ahí que la autoeficacia sea uno de los predictores individuales más consistentes del logro (Goetze y Driver, 2022). En contextos de lengua minorizada, donde la escuela ofrece pocas oportunidades de uso del inglés, cabe esperar que esas creencias cobren un peso aún mayor, junto con las emociones de aula (Guerra et al., 2024; Quispe et al., 2024). Se espera, por tanto, que la autoeficacia en inglés se asocie positivamente con el desempeño (H3).

18.2.2 Prácticas didácticas culturalmente pertinentes

La enseñanza culturalmente pertinente incorpora los saberes, las prácticas y la lengua de la comunidad como recurso del aprendizaje, en lugar de tratarlos como obstáculo. En la educación bilingüe rural latinoamericana, reconocer y valorar la diversidad lingüística y cultural de la comunidad es una condición del aprendizaje (Duque et al., 2024), y la enseñanza del inglés a estudiantes indígenas solo es aditiva —no erosiona la lengua propia— cuando se apoya en la alfabetización en la lengua originaria y en el castellano académico (Fabre et al., 2025). En el norte del Perú, los docentes de escuelas EIB consideran viable y prometedor integrar contenidos multimedia en lenguas originarias como recurso didáctico (Montoya et al., 2025). Desde la teoría social cognitiva, las tareas conectadas con la experiencia del estudiante aumentan la probabilidad de éxito temprano y ofrecen modelos cercanos, dos fuentes directas de autoeficacia. Se plantea que las prácticas didácticas culturalmente pertinentes se asocian positivamente con la autoeficacia en inglés (H1) y, a través de ella, con el desempeño (H4).

18.2.3 Barreras contextuales

La escuela rural andina enfrenta restricciones que no dependen del estudiante: falta de materiales y de conectividad, escaso tiempo de instrucción, cambios de docente durante el año y ausencia de recursos en la lengua originaria (Izquierdo et al., 2021; Guevara et al., 2026; Moreano et al., 2026). Estas barreras reducen las oportunidades de experiencia de dominio y, con ello, las creencias de eficacia; también pueden limitar el desempeño de forma directa, al restringir la exposición a la lengua. Se plantea que las barreras contextuales se asocian negativamente con la autoeficacia en inglés (H2) y, a través de ella, con el desempeño (H5). Las trayectorias directas de ambas condiciones de enseñanza sobre el desempeño se estiman como parte del modelo (mediación parcial).

18.2.4 Generalización entre comunidades de lengua originaria

Las lenguas quechua y aimara difieren en su estatus, su vitalidad y su presencia en los materiales escolares (Sullón, 2024), y el dominio del inglés de los estudiantes de secundaria varía de forma marcada entre tipos de escuela y regiones (Elizondo et al., 2026). Antes de afirmar que un mismo modelo describe a ambas comunidades, es necesario comprobar que los instrumentos miden lo mismo en las dos y que las trayectorias estructurales son equivalentes, condición para comparar medias y agregar resultados entre grupos. Se plantea que el modelo de medición y las trayectorias estructurales son invariantes entre estudiantes quechuahablantes y aimarahablantes (H6).

18.3 MATERIALES Y MÉTODOS

18.3.1 Diseño y participantes

Se realizó un estudio cuantitativo, explicativo y transversal, con un modelo de ecuaciones estructurales de variables latentes que especificó dos condiciones de enseñanza exógenas (prácticas didácticas culturalmente pertinentes, PDC, y barreras contextuales, BC), un mediador (autoeficacia en inglés, AE) y el desempeño en inglés (DES) medido con pruebas. El modelo hipotetizado fue de mediación parcial, con las trayectorias directas de ambas condiciones sobre el desempeño, y la comparación entre comunidades se abordó con una secuencia multigrupo de invarianza. Los datos se recogieron durante el año escolar 2025.

La población la constituyeron los estudiantes de 3.º a 5.º grado de secundaria de instituciones públicas rurales EIB de la sierra sur del Perú, con quechua o aimara como lengua materna e inglés como área obligatoria. El muestreo fue bietápico por conglomerados, estratificado por lengua originaria: se seleccionaron instituciones EIB rurales (Cusco, Apurímac, Ayacucho, Huancavelica y Puno en el estrato quechua; Puno en el aimara) y, en cada una, 15 estudiantes distribuidos entre los tres grados.

El tamaño muestral se fijó a priori con tres criterios, adoptándose el más exigente (Tabla 18-1). La prueba de ajuste cercano del RMSEA (MacCallum et al., 1996; $\varepsilon_0 = .05$, $\varepsilon_1 = .08$, $\alpha = .05$, gl = 84) alcanzó una potencia de .80 con N = 148. La potencia de las trayectorias y del efecto indirecto más pequeño se estimó por simulación de Monte Carlo (Donnelly et al., 2023; Moshagen y Bader, 2024), con 500 réplicas por tamaño y prueba de significación conjunta, fijando $b = .40$, coherente con r = .46 de la literatura, y $a = -.20$; el efecto indirecto resultante ($ab = -.08$) alcanzó una potencia de .806 con 250 casos efectivos. El efecto de diseño, $1 + (\bar{m} - 1)\cdot\rho_x\cdot\rho_y$ con $\bar{m} = 15$ y una correlación intraclase de planificación de .20, fue 1.56, de modo que la prueba de invarianza exigió 231 estudiantes por grupo. Redondeando a instituciones completas, el tamaño planificado y analizado fue N = 480 (16 instituciones × 15 estudiantes × 2 grupos; J = 32), con un tamaño efectivo de 307.7 y una potencia estimada por Monte Carlo de .854 con 300 casos efectivos. No hubo datos perdidos.

Tabla 18-1

Cálculo a priori del tamaño muestral por criterio

Criterio	Parámetros de entrada	Requisito efectivo	Ajuste por conglomerados (efecto de diseño = 1.56)	Requisito en estudiantes
Potencia de la prueba de ajuste cercano del RMSEA, modelo completo	$\varepsilon_0 = .05$; $\varepsilon_1 = .08$; gl = 84; $\alpha = .05$; $1 - \beta = .80$	148	— (se aplica por grupo)	148
Potencia del efecto indirecto más pequeño (BC → AE → DES)	$a = -.20$; $b = .40$; $ab = -.08$; significación conjunta; 500 réplicas por tamaño	250 (potencia = .806; con 225, .708)	250 × 1.56 = 390	390
Potencia de la trayectoria exógena más débil (BC → AE)	$\beta = -.20$; 500 réplicas por tamaño	250 (potencia = .806)	250 × 1.56 = 390	
Prueba de invarianza: cada grupo de lengua debe cumplir el requisito del RMSEA	$\varepsilon_0 = .05$; $\varepsilon_1 = .08$; gl = 84 por grupo	148 por grupo	148 × 1.56 = 230.9 → 231 por grupo	2 × 231 = 462

Criterio	Parámetros de entrada	Requisito efectivo	Ajuste por conglomerados (efecto de diseño = 1.56)	Requisito en estudiantes
Tamaño planificado y analizado	Requisito más exigente (462) redondeado a instituciones completas de 15 estudiantes	—	16 instituciones × 15 × 2 grupos	480

Nota: Potencia del RMSEA según MacCallum et al. (1996); potencia de las trayectorias y del efecto indirecto por simulación de Monte Carlo (Donnelly et al., 2023; Moshagen y Bader, 2024) con la prueba de significación conjunta; efecto de diseño para coeficientes de regresión con muestreo bietápico, $1 + (\bar{m} - 1)\rho\rho$, con $\bar{m} = 15$ y $\rho = .20$. Con N = 480 el tamaño efectivo es 480/1.56 = 307.7 y la potencia estimada por Monte Carlo del efecto indirecto más pequeño con 300 casos efectivos es .854.

18.3.2 Instrumentos

PDC y BC se midieron con cuatro ítems de frecuencia cada una (1 = nunca a 5 = siempre), sobre el uso de situaciones, lengua y cultura de la comunidad en la clase de inglés y sobre falta de materiales, suspensión de clases, rotación docente y ausencia de materiales en la lengua originaria, respectivamente. La AE se midió con cuatro ítems de acuerdo (1 a 5) sobre comprensión lectora, comprensión de instrucciones orales, escritura y conversación. El DES se midió con tres pruebas (comprensión lectora, comprensión auditiva y producción escrita), alineadas con el currículo nacional y calificadas en la escala vigesimal (0–20). Los indicadores figuran en la Tabla 18-3.

18.3.3 Análisis de datos

El análisis siguió dos pasos, modelo de medición y modelo estructural. Los indicadores de cinco categorías se trataron como continuos, opción de propiedades comparables a las de una variable continua cuando la escala tiene cinco o más puntos (Huh y Gim, 2025). Todos los modelos se estimaron por máxima verosimilitud en Python, con la escala de cada factor fijada por la carga de su primer indicador; las estimaciones del

modelo de un grupo coincidieron con las del paquete semopy. La fiabilidad se evaluó con alfa de Cronbach y omega de McDonald (Kalkbrenner, 2024); la validez convergente, con la varianza media extraída (VME ≥ .50); y la validez discriminante, con la raíz cuadrada de la VME frente a las correlaciones entre factores y con la razón heterorrasgo-monorrasgo (HTMT < .85), según Cheung et al. (2024). El ajuste se valoró con χ^2, CFI, TLI, RMSEA con su intervalo al 90 % y SRMR, con los valores de referencia de Hu y Bentler (1999): CFI cercano o superior a .95, RMSEA cercano o inferior a .06 y SRMR cercano o inferior a .08. El modelo hipotetizado se comparó con cuatro modelos rivales (un factor; tres factores con AE y DES fusionados; medición de cuatro factores; mediación completa).

Como los estudiantes están anidados en instituciones, que fueron la unidad de muestreo de la primera etapa y la situación en que corresponde ajustar la inferencia por conglomerados (Abadie et al., 2023), los errores estándar y los intervalos de las trayectorias se obtuvieron por bootstrap de conglomerados: 1 000 remuestras de instituciones con reposición dentro de cada grupo de lengua, con intervalo percentil al 95 %. Los efectos indirectos se estimaron como producto de las trayectorias estandarizadas con el mismo bootstrap (Preacher y Hayes, 2008). La dependencia por institución se cuantificó con la correlación intraclase CCI(1) de la puntuación media de cada constructo.

La invarianza entre quechuahablantes y aimarahablantes se evaluó con la secuencia configural, métrica, escalar y estructural (además, trayectorias iguales). Una restricción se consideró admisible cuando ΔCFI ≥ −.010 (Cheung y Rensvold, 2002) y ΔRMSEA ≤ .015 (Leitgöb et al., 2023). Sobre el modelo escalar, cada trayectoria se comparó entre grupos con una prueba $\Delta\chi^2(1)$, y los efectos indirectos de cada grupo se acompañaron de intervalos de Monte Carlo al 95 % (Preacher y Selig, 2012); en los modelos multigrupo, los errores estándar son los de máxima verosimilitud. El nivel de significación fue $\alpha = .05$.

18.4 RESULTADOS

18.4.1 Características de la muestra

La muestra analizada estuvo formada por 480 estudiantes de 32 instituciones educativas, 240 quechuahablantes y 240 aimarahablantes (Tabla 18-2). Los dos grupos no difirieron en sexo, grado, edad ni frecuencia de uso de la lengua originaria en el hogar; sí difirieron en la carga horaria semanal de inglés, mayor en el grupo quechua, característica de la institución y no del estudiante.

Tabla 18-2

Características de la muestra por lengua originaria (N = 480 estudiantes; J = 32 instituciones)

Característica	Quechua	Aimara	Total	Contraste entre grupos
Estudiantes, *n*	240	240	480	
Instituciones educativas, *J*	16	16	32	
Estudiantes por institución, *M* (*DE*)	15.0 (0.0)	15.0 (0.0)	15.0 (0.0)	—
Mujeres, %	52.5	53.8	53.1	$\chi^2(1) = 0.08$; $p = .784$; $V = .01$
3.er grado de secundaria, %	29.2	32.9	31.0	$\chi^2(2) = 0.79$; $p = .672$; $V = .04$
4.º grado de secundaria, %	36.7	35.0	35.8	—
5.º grado de secundaria, %	34.2	32.1	33.1	—
Edad en años, *M* (*DE*)	15.1 (1.0)	15.0 (1.1)	15.1 (1.0)	*t* de Welch = 1.09; $p = .276$
Uso diario de la lengua originaria en el hogar, %	62.9	57.1	60.0	$\chi^2(2) = 1.87$; $p = .393$; $V = .06$
Uso frecuente, %	24.2	29.2	26.7	—
Uso ocasional, %	12.9	13.8	13.3	—
Dos horas semanales de inglés, %	62.5	81.2	71.9	$\chi^2(1) = 20.87$; $p < .001$; $V = .21$
Tres horas semanales de inglés, %	37.5	18.8	28.1	—

Nota: Los porcentajes se calculan sobre el total de estudiantes de cada grupo. La carga horaria semanal de inglés es una característica de la institución. V = V de Cramér.

Tabla 18-3

Indicadores, estadísticos descriptivos y cargas factoriales estandarizadas (N = 480)

Constructo	Indicador		M	DE	Asimetría	Curtosis	λ	R^2
PDC	PDC1	El docente relaciona las actividades de inglés con situaciones de mi comunidad	3.24	1.11	−0.30	−0.62	.82	.68
	PDC2	Usamos palabras o ejemplos de mi lengua originaria para entender el inglés	3.21	1.11	−0.25	−0.61	.72	.52
	PDC3	Las tareas de inglés incluyen temas de la cultura andina (fiestas, trabajo agrícola, saberes locales)	3.25	1.13	−0.28	−0.66	.79	.63
	PDC4	El docente valora que usemos el quechua o el aimara durante la clase de inglés	3.26	1.11	−0.19	−0.64	.69	.47
BC	BC1	En mi colegio faltan materiales para aprender inglés (libros, audios, diccionarios)	3.53	1.08	−0.35	−0.52	.72	.52
	BC2	Las clases de inglés se suspenden o se acortan por distancia, clima u otras actividades	3.49	1.07	−0.37	−0.45	.64	.41
	BC3	Durante el año cambiamos de docente de inglés o pasamos semanas sin docente	3.54	1.08	−0.41	−0.46	.62	.39
	BC4	No hay materiales de inglés que usen mi lengua originaria o mi cultura	3.54	1.13	−0.35	−0.73	.77	.60
AE	AE1	Puedo entender lo principal de un texto corto en inglés	3.01	1.06	−0.02	−0.66	.79	.63
	AE2	Puedo comprender las instrucciones sencillas que el docente da en inglés	2.99	1.06	−0.02	−0.52	.73	.54
	AE3	Puedo escribir un párrafo corto en inglés sobre mi vida diaria	2.96	1.12	−0.11	−0.75	.73	.54
	AE4	Puedo presentarme y conversar brevemente en inglés con un compañero	2.97	1.10	−0.04	−0.57	.74	.55
DES	DES_L	Prueba de comprensión lectora (0–20)	11.04	2.98	0.11	0.23	.79	.62
	DES_A:	Prueba de comprensión auditiva (0–20)	11.18	3.12	−0.15	−0.07	.75	.56
	DES_E	Prueba de producción escrita (0–20)	11.22	3.10	0.01	0.05	.65	.42

Nota: PDC y BC: escala de frecuencia de 1 (nunca) a 5 (siempre); AE: escala de acuerdo de 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo); DES: puntuaciones de pruebas en la escala vigesimal (0–20). λ = carga estandarizada del modelo de medición de cuatro factores; curtosis = exceso de curtosis.

18.4.2 Modelo de medición

Los indicadores mostraron asimetrías y curtosis moderadas (valor absoluto máximo de asimetría = 0.41 y de exceso de curtosis = 0.75), y todas las cargas estandarizadas del modelo de cuatro factores superaron .60 (rango .62–.82; Tabla 18-3).

La fiabilidad fue de $\alpha = .77–.84$ y $\omega = .77–.84$ en los cuatro constructos (Tabla 18-4). La varianza media extraída superó .50 en prácticas, autoeficacia y desempeño, y fue de .48 en las barreras contextuales, cuya carga más baja correspondió al indicador de rotación o ausencia docente (BC3, $\lambda = .62$); el indicador se conservó por su contenido y la fiabilidad del constructo se mantuvo ($\omega = .79$). La dependencia por institución fue mayor en las barreras (CCI = .214) que en el desempeño (.188), las prácticas (.178) y la autoeficacia (.054).

Tabla 18-4

Fiabilidad, validez convergente y dependencia por institución de los constructos

Constructo	Indicadores	α	ω	VME	λ mín.–máx.	CCI(1)	Efecto de diseño de la media
Prácticas didácticas culturalmente pertinentes (PDC)	4	.84	.84	.57	.69–.82	.178	3.49
Barreras contextuales (BC)	4	.79	.79	.48	.62–.77	.214	4.00
Autoeficacia en inglés (AE))	4	.84	.84	.57	.73–.79	.054	1.75
Desempeño en inglés (DES)	3	.77	.77	.53	.65–.79	.188	3.63

Nota: α = alfa de Cronbach; ω = omega de McDonald calculado con las cargas estandarizadas; VME = varianza media extraída; CCI(1) = correlación intraclase de la puntuación media del constructo entre instituciones (J = 32; tamaño medio de conglomerado = 15.0); efecto de diseño de la media = $1 + (\bar{n}_0 - 1)\cdot CCI$.

La raíz cuadrada de la VME de cada constructo superó sus correlaciones con los demás, y todas las razones HTMT fueron inferiores a .85 (máximo = .48, entre AE y DES; Tabla 18-5).

Tabla 18-5

Validez discriminante: raíz cuadrada de la VME frente a las correlaciones entre factores y razón heterorrasgo-monorrasgo (HTMT)

Constructo	PDC	BC	AE	DES
Prácticas didácticas culturalmente pertinentes (PDC)	.76	[.32]	[.38]	[.30]
Barreras contextuales (BC)	−.33	.69	[.22]	[.22]
Autoeficacia en inglés (AE))	.37	−.22	.75	[.48]
Desempeño en inglés (DES)	.31	−.22	.46	.73

Nota: Diagonal (negrita): raíz cuadrada de la VME. Debajo de la diagonal: correlaciones entre factores del modelo de medición. Encima de la diagonal, entre corchetes: HTMT (formulación original con medias aritméticas de las correlaciones entre indicadores). PDC = prácticas didácticas culturalmente pertinentes; BC = barreras contextuales; AE = autoeficacia en inglés; DES = desempeño en inglés.

18.4.3 Ajuste de los modelos

El modelo de un factor (M1) y el de tres factores con autoeficacia y desempeño fusionados (M2) ajustaron peor que el de cuatro factores (M3); la fusión de autoeficacia y desempeño empeoró significativamente el ajuste ($\Delta\chi^2(3) = 271.63$, $p < .001$; Tabla 18-6). El modelo hipotetizado de mediación parcial (M5) ajustó a los datos ($\chi^2(84) = 108.16$, $p = .039$; CFI = .991; TLI = .988; RMSEA = .024, IC 90 % [.006, .037]; SRMR = .034), dentro de los valores de referencia. Restringir a cero las dos trayectorias directas (M4, mediación completa) empeoró el ajuste ($\Delta\chi^2(2) = 10.47$, $p = .005$), y se retuvo el modelo de mediación parcial.

18.4.4 Trayectorias estructurales (H1–H3)

Las prácticas didácticas culturalmente pertinentes se asociaron positivamente con la autoeficacia (β = .335, IC 95 % [.232, .435]; H1 se sostiene). La trayectoria de las barreras contextuales hacia la autoeficacia fue negativa pero no se distinguió de cero (β = −.115, IC 95 % [−.239,

.003]; $z = -1.79$, $p = .074$; H2 no se sostiene); su magnitud fue menor que el efecto mínimo de interés fijado a priori ($\beta = -.20$), valor que el intervalo no excluye. La autoeficacia se asoció positivamente con el desempeño ($\beta = .393$, IC 95 % [.285, .494]; H3 se sostiene).

Tabla 18-6

Ajuste de los modelos de medición y estructurales rivales (N = 480)

	Modelo	χ^2	gl	p	CFI	TLI	RMSEA [IC 90 %	SRMR	AIC	BIC
M!	Un factor	1410.67	90	< .001	.486	.400	.175 [.167, .183]	.141	1470.7	1595.9
M2	Tres factores (AE y DES fusionados)	379.80	87	< .001	.886	.862	.084 [.075, .092]	.068	445.8	583.5
M3	Medición, cuatro factores correlacionados	108.16	84	.039	.991	.988	.024 [.006, .037]	.034	180.2	330.4
M4	Estructural, mediación completa	118.63	86	.011	.987	.984	.028 [.014, .040]	.044	186.6	328.5
M5	Estructural, mediación parcial (hipotetizado)	108.16	84	.039	.991	.988	.024 [.006, .037]	.034	180.2	330.4

Nota: Estimación por máxima verosimilitud. M5 tiene una parte estructural saturada y su ajuste es, por tanto, idéntico al de M3. Comparación M2 frente a M3: $\Delta\chi^2(3) = 271.63$, $p < .001$. Comparación M4 frente a M5: $\Delta\chi^2(2) = 10.47$, $p = .005$. Valores de referencia: CFI cercano o superior a .95, RMSEA cercano o inferior a .06 y SRMR cercano o inferior a .08 (Hu y Bentler, 1999).

Ninguna de las dos trayectorias directas sobre el desempeño se distinguió de cero por separado (prácticas: $\beta = .130$, IC 95 % [−.003, .254]; barreras: $\beta = -.094$, IC 95 % [−.208, .019]). El modelo explicó el 15.1 % de la varianza de la autoeficacia y el 24.3 % de la del desempeño (Tabla 18-7; Figura 18-1).

Tabla 18-7

Trayectorias estructurales del modelo hipotetizado (M5)

Hipótesis	Trayectoria	*B*	*EE*	β	*z*	*p*	IC 95 % de β	Decisión
H1	PDC → AE	0.308	0.054	.335	5.68	< .001	[.232, .435]	Se sostiene
H2	BC → AE	−0.125	0.070	−.115	−1.79	.074	[−.239, .003]	No se sostiene
H3	AE → DES	1.096	0.184	.393	5.95	< .001	[.285, .494]	Se sostiene
—	PDC → DES (directa)	0.333	0.177	.130	1.89	.059	[−.003, .254]	No distinta de cero
—	BC → DES (directa)	−0.284	0.186	−.094	−1.53	.127	[−.208, .019]	No distinta de cero
—	R^2 de AE	—	—	.151	—	—	[.091, .236]	—
—	R^2 de DES	—	—	.243	—	—	[.163, .346]	—

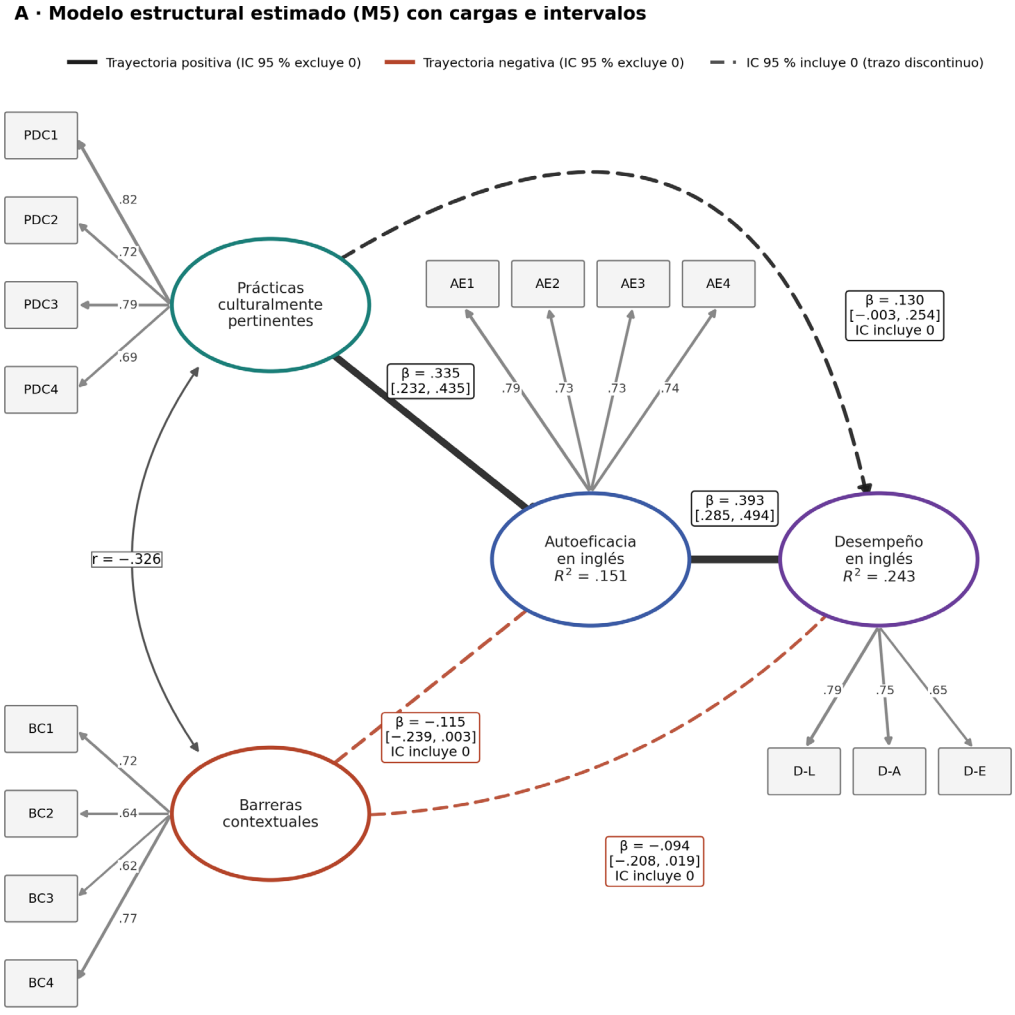
Nota: B = coeficiente no estandarizado; EE = error estándar por bootstrap de conglomerados (1000 remuestras de instituciones dentro de cada grupo de lengua); β = coeficiente estandarizado; IC 95 % percentil del bootstrap de conglomerados. z = B/EE.

18.4.2 Modelo de medición

Los indicadores mostraron asimetrías y curtosis moderadas (valor absoluto máximo de asimetría = 0.41 y de exceso de curtosis = 0.75), y todas las cargas estandarizadas del modelo de cuatro factores superaron .60 (rango .62–.82; Tabla 18-3).

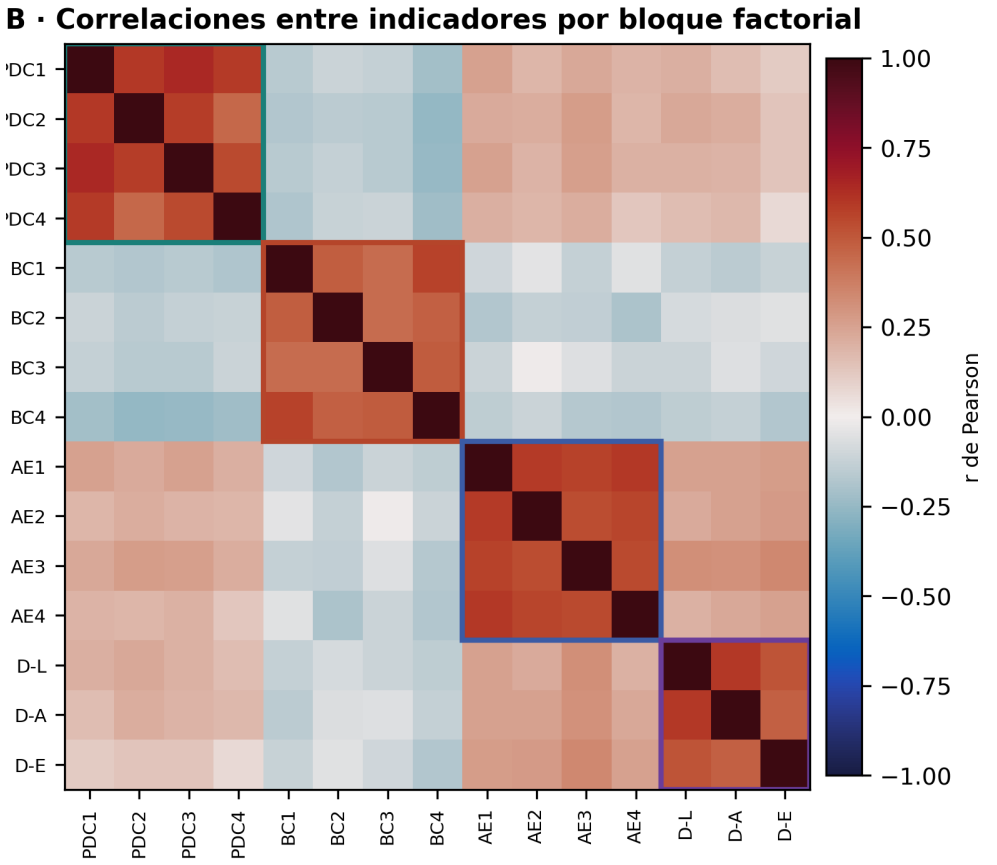
Figura 18-1

Modelo estructural estimado, correlaciones entre indicadores y validez discriminante

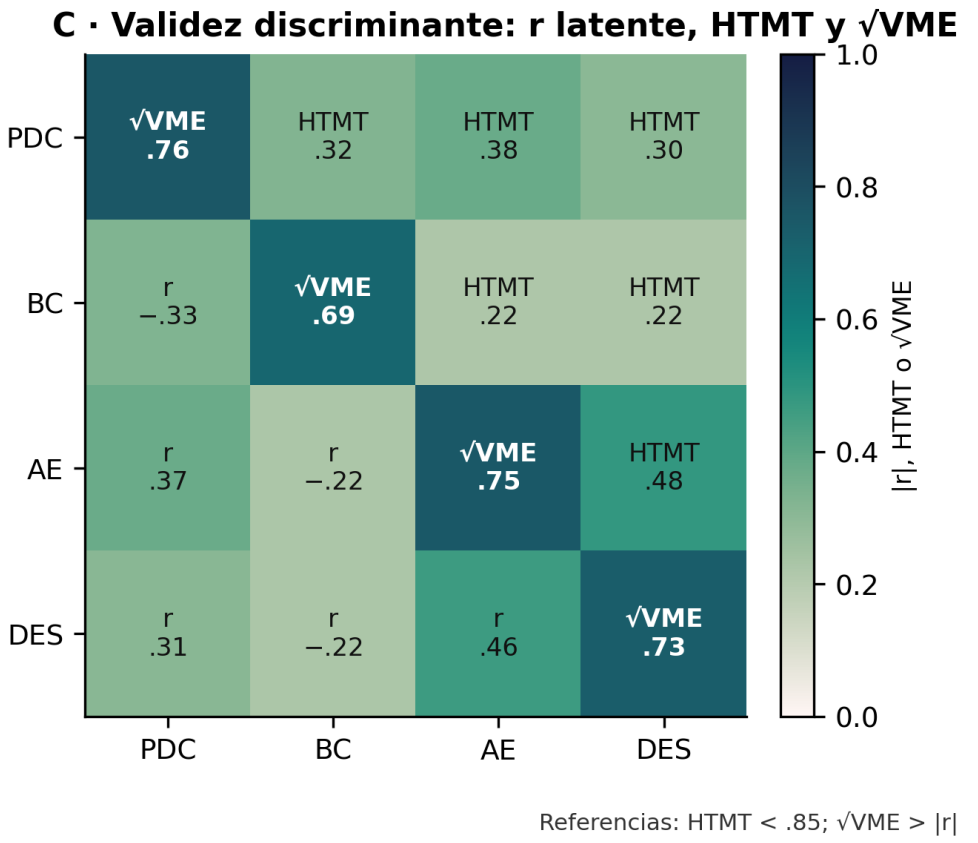


$\chi^2(84) = 108.16 \cdot CFI = .991 \cdot TLI = .988 \cdot RMSEA = .024\ [.006, .037] \cdot SRMR = .034 \cdot N = 480$

Nota. Modelo de ecuaciones estructurales estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones). Panel A: trayectorias estandarizadas con IC 95 % del bootstrap de conglomerados (1000 remuestras); el grosor de cada flecha es proporcional a |β| y el trazo discontinuo indica que el intervalo incluye el cero: PDC → AE β = .335 [.232, .435]; BC → AE β = −.115 [−.239, .003]; AE → DES β = .393 [.285, .494]; PDC → DES β = .130 [−.003, .254]; BC → DES β = −.094 [−.208, .019]; correlación entre las condiciones exógenas r = −.326. Los números junto a los indicadores son cargas estandarizadas (rango .62–.82). R^2 de AE = .151 y de DES = .243. Ajuste: $\chi^2(84) = 108.16$; CFI = .991; TLI = .988; RMSEA = .024 [IC 90 % .006, .037]; SRMR = .034.



Nota. Modelo de ecuaciones estructurales estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones). Panel B: correlaciones de Pearson entre los 15 indicadores; los recuadros delimitan los bloques de cada factor.



Nota. Modelo de ecuaciones estructurales estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones). Panel C: debajo de la diagonal, correlaciones entre factores del modelo de medición; encima, razón HTMT (máximo = .48); en la diagonal, raíz cuadrada de la VME (.69 a .76). PDC = prácticas didácticas culturalmente pertinentes; BC = barreras contextuales; AE = autoeficacia en inglés; DES = desempeño en inglés; D-L, D-A y D-E = pruebas de comprensión lectora, comprensión auditiva y producción escrita.

18.4.5 Efectos indirectos (H4–H5)

El efecto indirecto de las prácticas culturalmente pertinentes sobre el desempeño a través de la autoeficacia fue .132 (IC 95 % [.077, .195]; H4 se sostiene) y representó el 50.3 % de su efecto total (.261; IC 95 % [.143, .373]). El efecto indirecto de las barreras fue –.045 (IC 95 % [–.098, .001]; H5 no se sostiene), mientras que su efecto total fue negativo y distinto de cero (–.139; IC 95 % [–.258, –.034]). El contraste entre la magnitud de ambos efectos indirectos fue .087 (IC 95 % [.008, .172]; Tabla 18-8; Figura 18-2).

Tabla 10-8

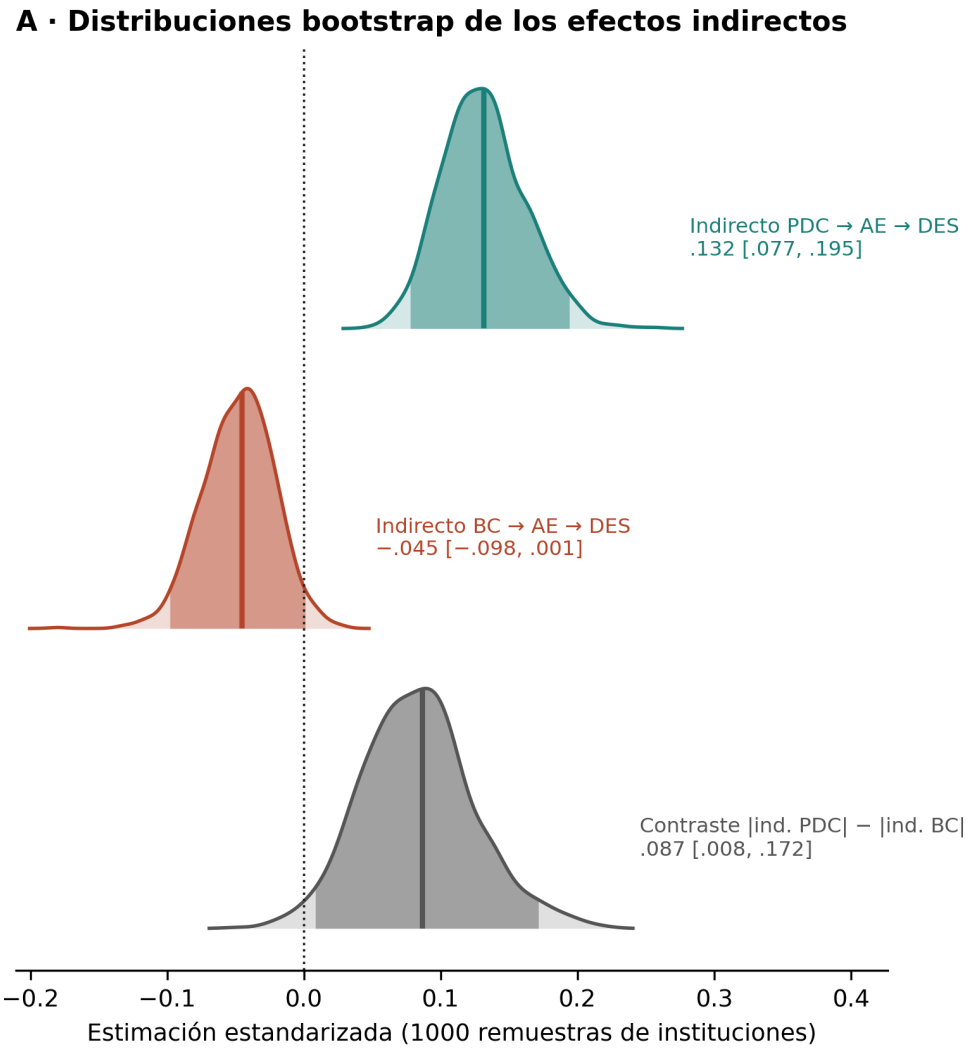
Descomposición de efectos del aprendizaje autorregulado sobre la ansiedad matemática

Condición de enseñanza	Efecto	Estimación	*EE*	IC 95 %	Hipótesis	Decisión
Prácticas didácticas culturalmente pertinentes	Indirecto vía AE	.132	.030	[.077, .195]	H4	Se sostiene
	Directo	.130	.066	[–.003, .254]	—	No distinto de cero
	Total	.261	.060	[.143, .373]	—	Distinto de cero
Barreras contextuales	Indirecto vía AE	–.045	.026	[–.098, .001]	H5	No se sostiene
	Directo	–.094	.060	[–.208, .019]	—	No distinto de cero
	Total	–.139	.056	[–.258, –.034]	—	Distinto de cero
Contraste	|indirecto PDC| – |indirecto BC|	.087	.041	[.008, .172]	—	Distinto de cero

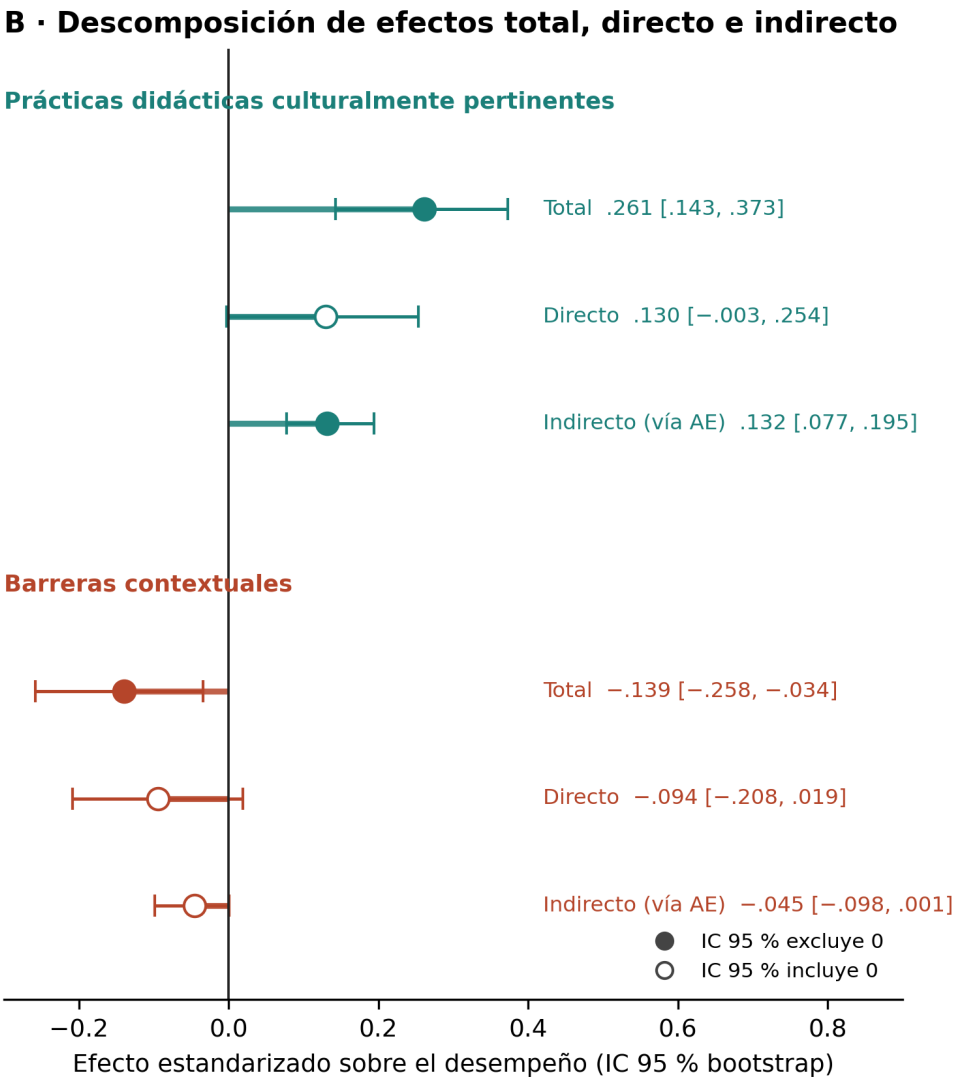
Nota: Efecto indirecto = producto de las trayectorias estandarizadas a·b (Preacher y Hayes, 2008); EE e IC 95 % percentil por bootstrap de conglomerados (1000 remuestras). Proporción del efecto total de PDC transmitida por la autoeficacia = .50; para BC no se calcula porque su efecto indirecto no se distingue de cero.

Figura 18-2

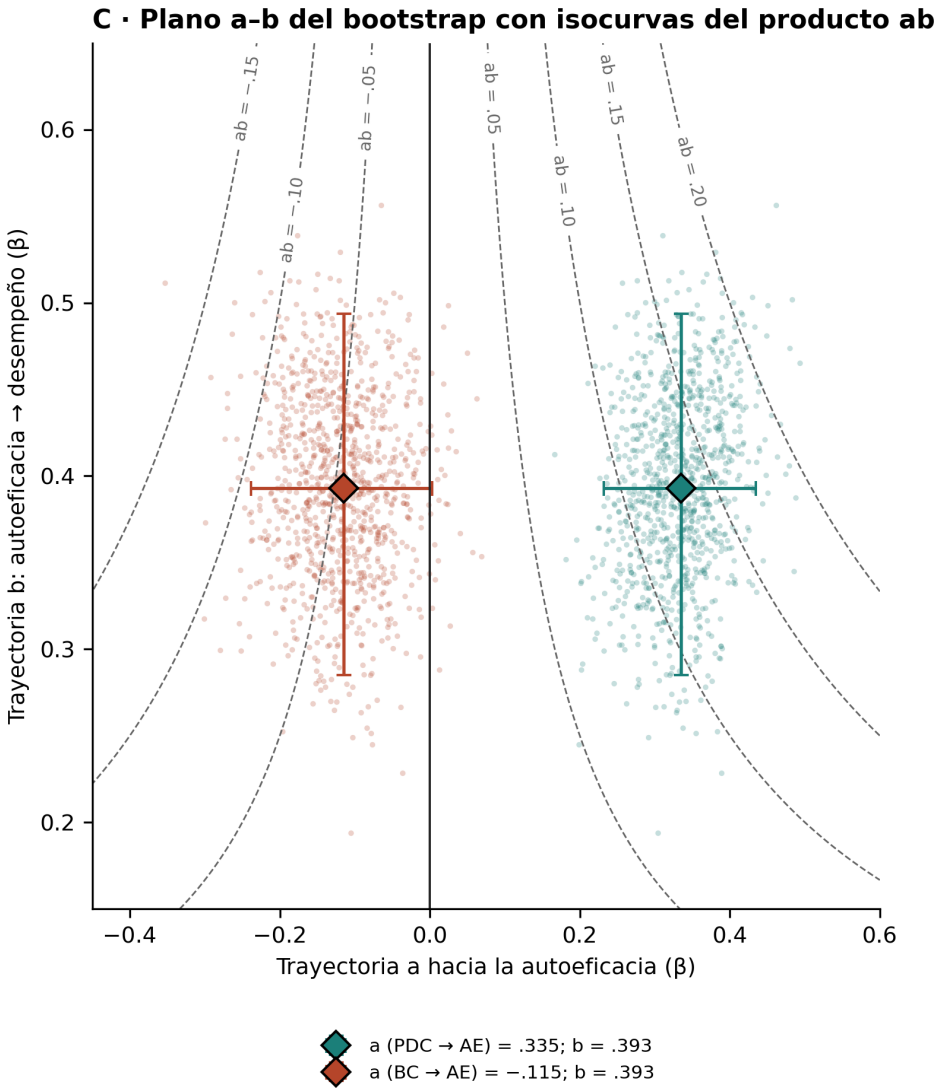
Distribuciones bootstrap, descomposición de efectos y plano de las trayectorias de mediación



Nota. Coeficientes estandarizados del modelo hipotetizado estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones); 1000 remuestras de instituciones con reposición dentro de cada grupo de lengua. Panel A: densidad de las estimaciones bootstrap del efecto indirecto de PDC (.132 [.077, .195]), del efecto indirecto de BC (−.045 [−.098, .001]) y del contraste entre sus magnitudes (.087 [.008, .172]); la zona sombreada intensa es el IC 95 % percentil y la línea vertical, la estimación puntual.



Nota. Coeficientes estandarizados del modelo hipotetizado estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones); 1000 remuestras de instituciones con reposición dentro de cada grupo de lengua. Panel B: efectos total, directo e indirecto (vía AE) de cada condición de enseñanza sobre el desempeño, con IC 95 %; el círculo relleno indica que el intervalo excluye el cero. PDC: total .261, directo .130, indirecto .132. BC: total −.139, directo −.094, indirecto −.045.



Nota. Coeficientes estandarizados del modelo hipotetizado estimado por máxima verosimilitud (N = 480 estudiantes; J = 32 instituciones); 1000 remuestras de instituciones con reposición dentro de cada grupo de lengua. Panel C: cada punto es una remuestra en el plano formado por la trayectoria a (hacia la autoeficacia) y la trayectoria b (autoeficacia → desempeño); las líneas discontinuas son isocurvas del producto ab y los rombos, las estimaciones con su IC 95 %.

18.4.6 Invarianza entre comunidades de lengua originaria (H6)

El modelo configural ajustó en ambos grupos (CFI = .992; RMSEA = .022). Las restricciones de igualdad de cargas (ΔCFI = .001; ΔRMSEA = −.003), de interceptos (ΔCFI = .002; ΔRMSEA = −.004) y de trayectorias estructurales (ΔCFI = −.001; ΔRMSEA = .001) no empeoraron el ajuste según los criterios establecidos, y H6 se sostiene (Tabla 18-9).

Tabla 18-9

Invarianza de la medición y de las trayectorias estructurales entre estudiantes quechuahablantes (n = 240) y aimarahablantes (n = 240)

Nivel	χ^2	gl	CFI	RMSEA [IC 90 %]	SRMR	Comparación	$\Delta\chi^2$ (Δgl)	p
Configural	188.37	168	.992	.022 [.000, .038]	.042	—	—	—
Métrica (cargas iguales)	195.81	179	.993	.020 [.000, .036]	.045	frente a Configural	7.43 (11)	.763
Escalar (cargas e interceptos iguales)	201.41	190	.996	.016 [.000, .033]	.043	frente a Métrica	5.60 (11)	.899
Estructural (además, trayectorias iguales)	208.43	195	.995	.017 [.000, .033]	.051	frente a Escalar	7.02 (5)	.219

Nivel	ΔCFI	ΔRMSEA	Decisión
Configural	—	—	Línea base
Métrica (cargas iguales)	.001	−.003	Invarianza sostenida
Escalar (cargas e interceptos iguales)	.002	−.004	Invarianza sostenida
Estructural (además, trayectorias iguales)	−.001	.001	Invarianza sostenida

Nota: Modelos multigrupo estimados por máxima verosimilitud sobre el modelo estructural hipotetizado. Criterios de no invarianza: ΔCFI < −.010 (Cheung y Rensvold, 2002) y ΔRMSEA > .015 (Leitgöb et al., 2023). Diferencias de medias latentes del grupo aimara respecto del quechua en el modelo escalar (escala del indicador de referencia; z, p): PDC = −0.126 (EE = 0.091; z = −1.39; p = .165); BC = 0.208 (EE = 0.080; z = 2.61; p = .009); AE = −0.089 (EE = 0.081; z = −1.10; p = .271); DES = 0.097 (EE = 0.230; z = 0.42; p = .674).

Con invarianza escalar, la única diferencia de medias latentes entre grupos fue la de las barreras contextuales, más altas en el grupo aimara (diferencia = 0.208 en la escala del indicador de referencia; $z = 2.61$, $p = .009$).

Ninguna trayectoria difirió entre grupos en las pruebas de igualdad una a una (todas las $\Delta\chi^2(1)$ con $p \geq .181$). El efecto indirecto de las prácticas fue positivo y distinto de cero en ambos grupos, mientras que el de las barreras no se distinguió de cero en ninguno (Tabla 18-10; Figura 18-3).

Tabla 18-10

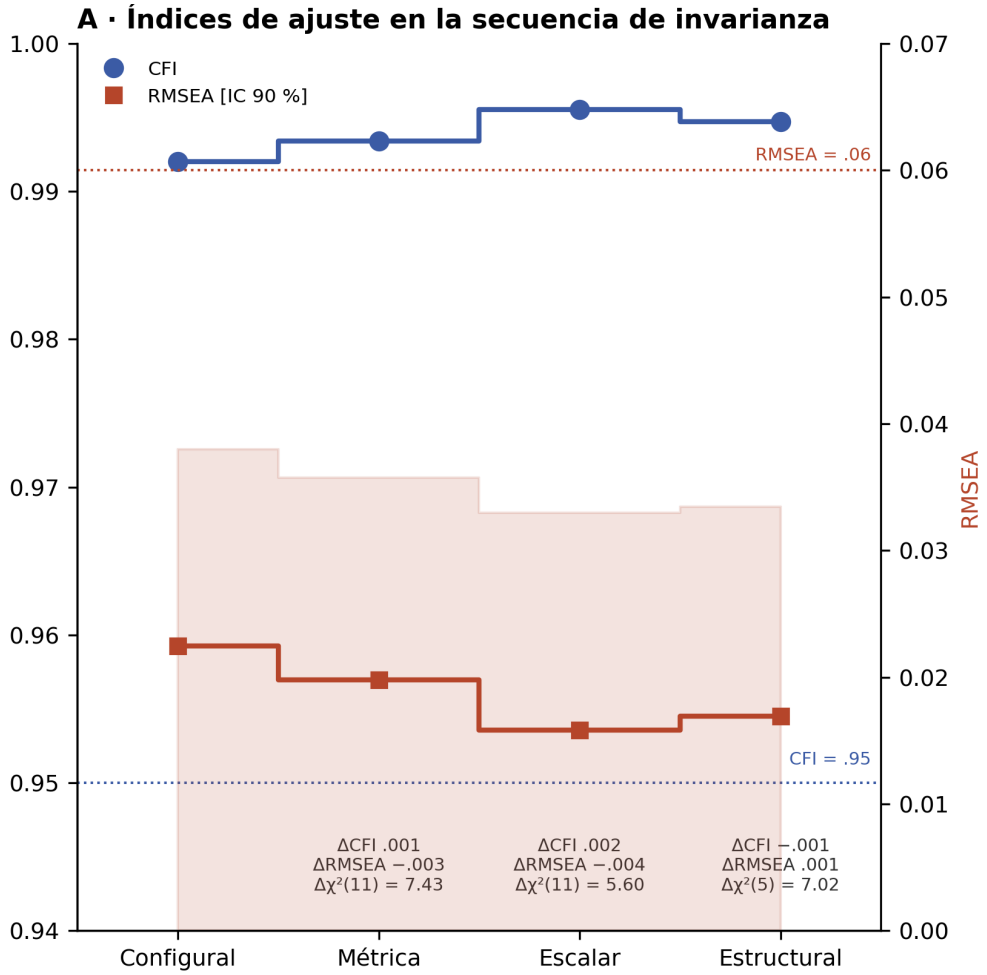
Comparación de las trayectorias estructurales y de los efectos indirectos entre grupos de lengua originaria (modelo con invarianza escalar)

Trayectoria	Quechua		Aimara		Δβ (Aimara – Quechua)	Δχ²(1) por igualdad	*p*
	β	*B* (*EE*)	β	*B* (*EE*)			
PDC → AE	.302	0.282 (0.073)	.371	0.333 (0.073)	.070	0.25	.616
BC → AE	–.148	–0.161 (0.087)	–.067	–0.072 (0.088)	.081	0.52	.473
AE → DES	.361	0.944 (0.222)	.421	1.249 (0.244)	.060	0.90	.342
PDC → DES (directa)	.043	0.106 (0.209)	.193	0.512 (0.222)	.149	1.79	.181
BC → DES (directa)	–.066	–0.188 (0.239)	–.122	–0.392 (0.251)	–.056	0.35	.556
Indirecto PDC → AE → DES	.109 [.044, .192]	—	.156 [.075, .252]	—	.047	—	—
Indirecto BC → AE → DES	–.053 [–.123, .002]	—	–.028 [–.099, .040]	—	.025	—	—

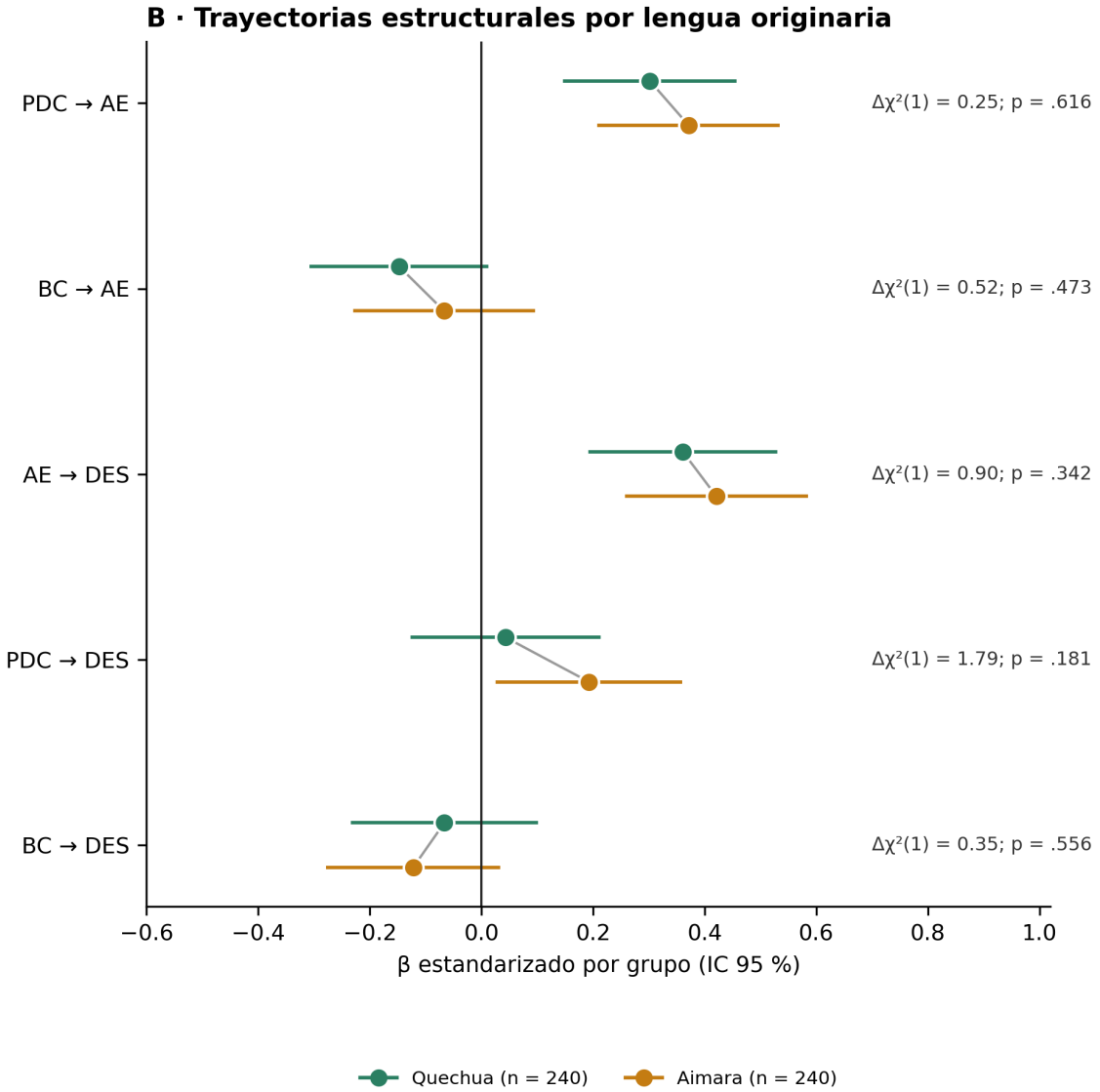
Nota: Cada prueba $\Delta\chi^2(1)$ compara el modelo con invarianza escalar y trayectorias libres con el mismo modelo en el que solo esa trayectoria se iguala entre grupos. EE de máxima verosimilitud. Efectos indirectos por grupo con intervalo de Monte Carlo al 95 % (20 000 extracciones; Preacher y Selig, 2012).

Figura 18-3

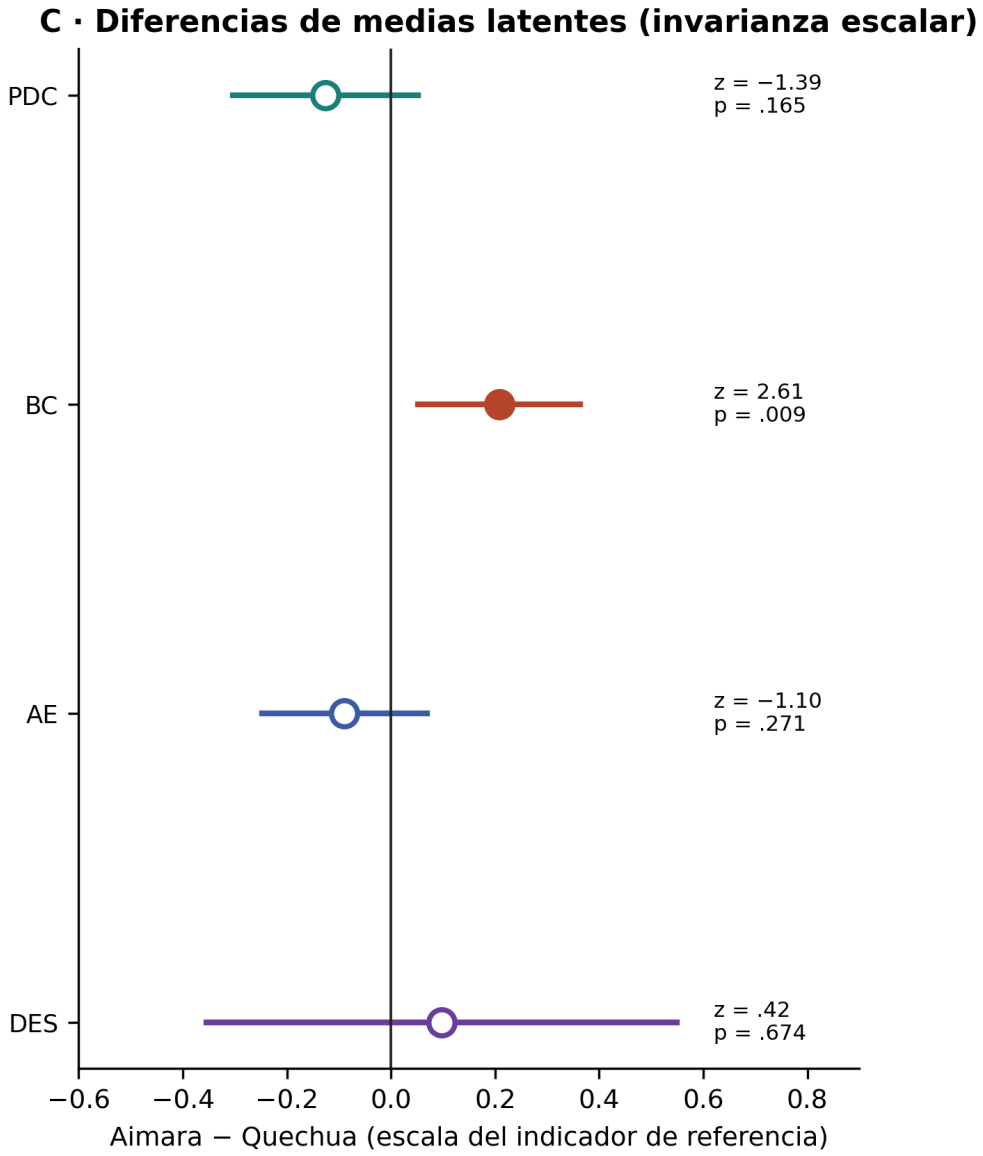
Invarianza de la medición, trayectorias estructurales por grupo y medias latentes entre comunidades de lengua originaria



Nota. Modelos multigrupo de máxima verosimilitud con estudiantes quechuahablantes (n = 240) y aimarahablantes (n = 240). Panel A: CFI (eje izquierdo) y RMSEA con su IC 90 % (eje derecho) en los modelos configural (CFI = .992; RMSEA = .022), métrico (ΔCFI = .001; ΔRMSEA = −.003), escalar (ΔCFI = .002; ΔRMSEA = −.004) y estructural (ΔCFI = −.001; ΔRMSEA = .001); las líneas punteadas marcan CFI = .95 y RMSEA = .06 (Hu y Bentler, 1999).



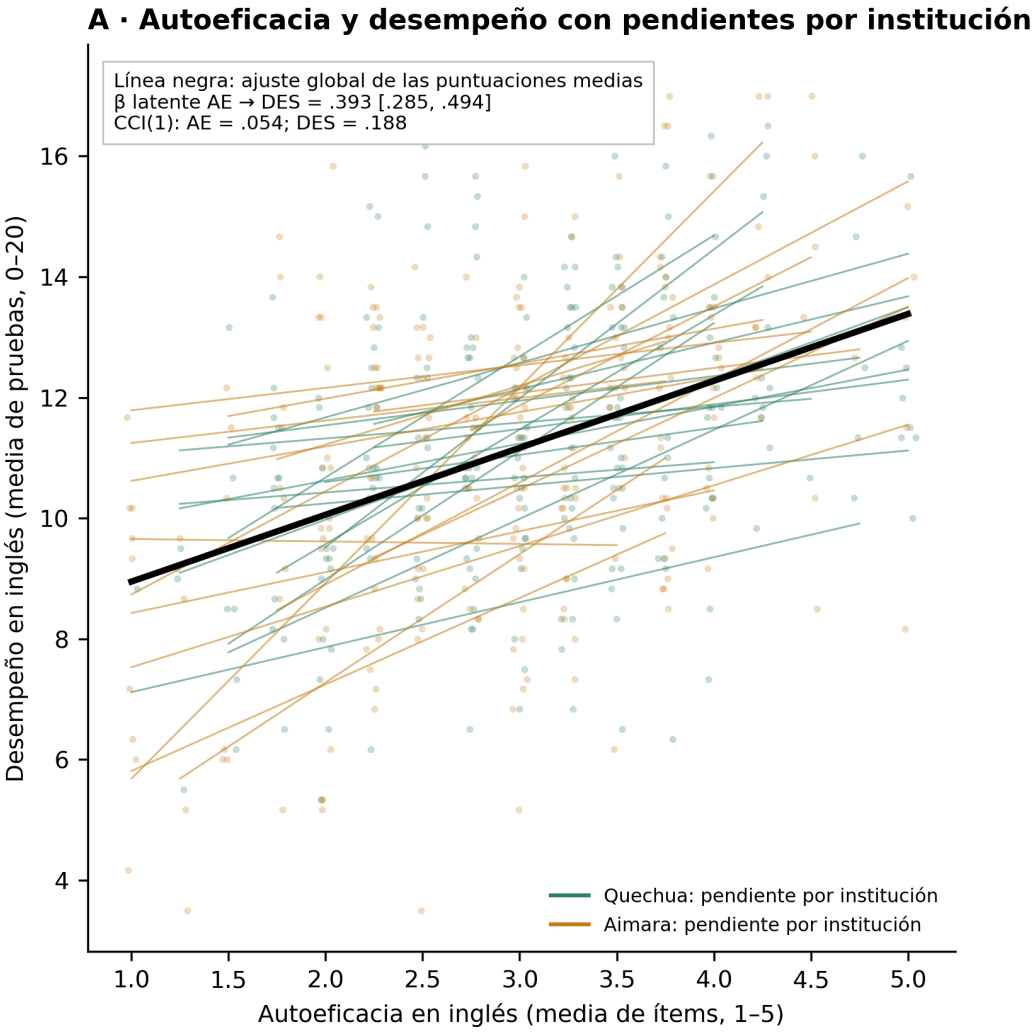
Nota. Modelos multigrupo de máxima verosimilitud con estudiantes quechuahablantes (n = 240) y aimarahablantes (n = 240). Panel B: trayectorias estandarizadas de cada grupo con IC 95 % de máxima verosimilitud y prueba de igualdad de cada trayectoria en el modelo escalar; ninguna difiere entre grupos (AE ← PDC: Quechua β = .302, Aimara β = .371, $\Delta\chi^2(1) = 0.25$, p = .616; AE ← BC: Quechua β = −.148, Aimara β = −.067, $\Delta\chi^2(1) = 0.52$, p = .473; DES ← AE: Quechua β = .361, Aimara β = .421, $\Delta\chi^2(1) = 0.90$, p = .342; DES ← PDC: Quechua β = .043, Aimara β = .193, $\Delta\chi^2(1) = 1.79$, p = .181; DES ← BC: Quechua β = −.066, Aimara β = −.122, $\Delta\chi^2(1) = 0.35$, p = .556).



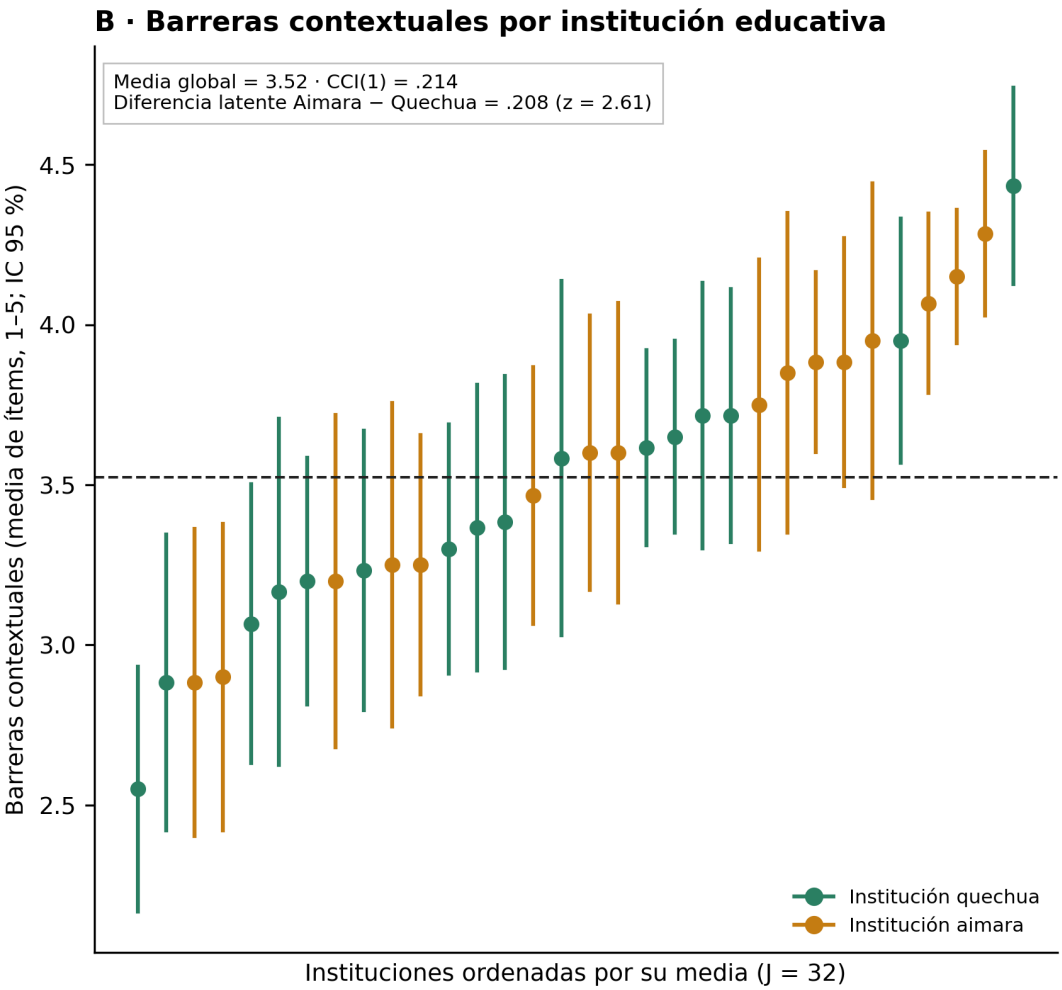
Nota. Modelos multigrupo de máxima verosimilitud con estudiantes quechuahablantes (n = 240) y aimarahablantes (n = 240). Panel C: diferencias de medias latentes Aimara − Quechua en el modelo con invarianza escalar, en la escala del indicador de referencia, con IC 95 %; el círculo relleno indica que el intervalo excluye el cero (BC = .208, z = 2.61, p = .009).

Figura 18-4

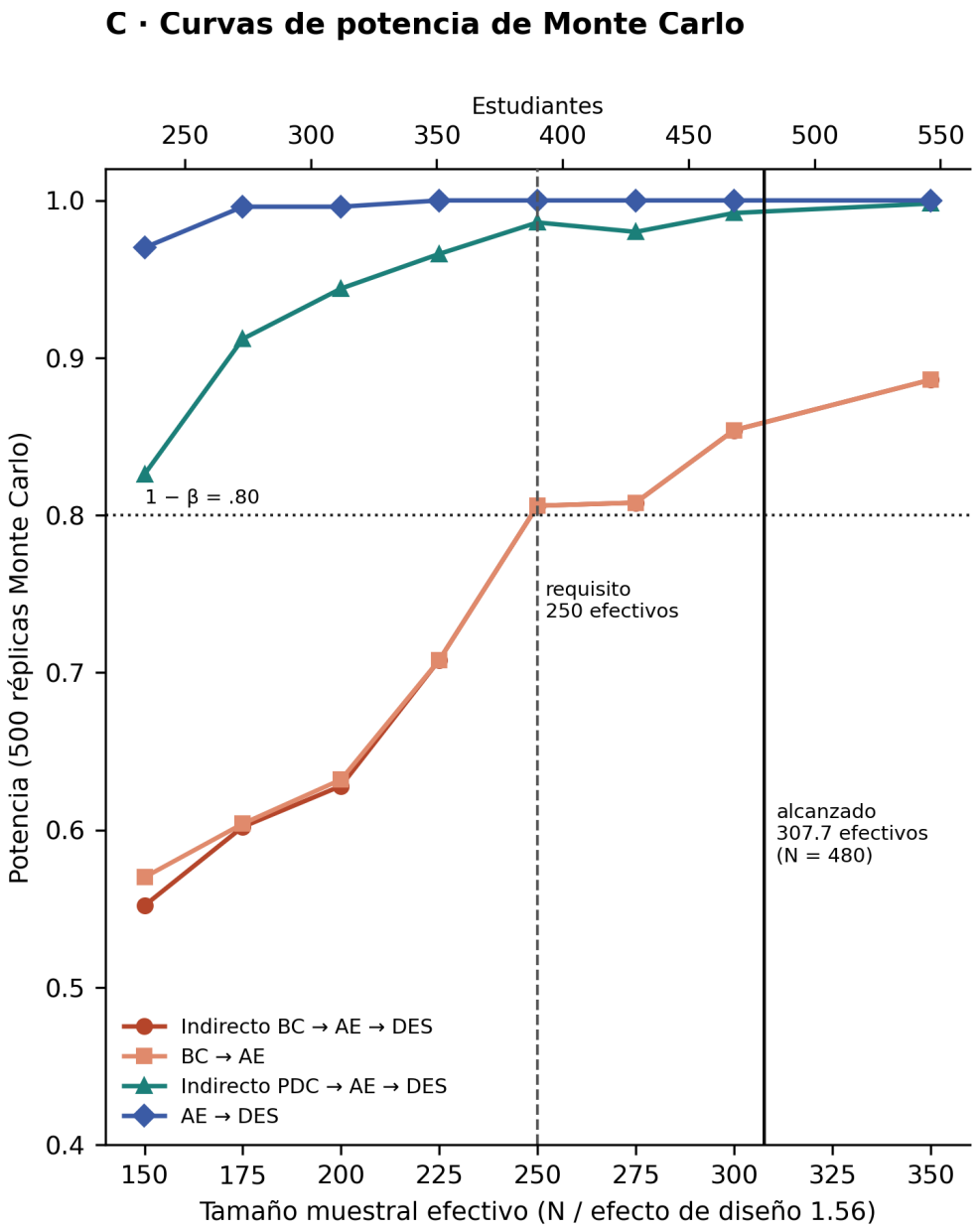
Relación entre autoeficacia y desempeño por institución, barreras contextuales por institución y curvas de potencia del diseño.



Nota. Datos de los 480 estudiantes de 32 instituciones; trayectoria latente del modelo estimado por máxima verosimilitud y potencia por simulación de Monte Carlo. Panel A: puntuaciones medias de autoeficacia (1–5) y de desempeño (0–20) de los 480 estudiantes, con la recta de mínimos cuadrados de cada institución (color por lengua originaria) y la recta global (negro); la trayectoria latente estimada es β = .393 [.285, .494]; CCI(1) de AE = .054 y de DES = .188..



Nota. Datos de los 480 estudiantes de 32 instituciones; trayectoria latente del modelo estimado por máxima verosimilitud y potencia por simulación de Monte Carlo. Panel B: media de las barreras contextuales de cada una de las 32 instituciones con IC 95 % (t), ordenadas de menor a mayor; la línea discontinua es la media global; CCI(1) = .214.



Nota. Datos de los 480 estudiantes de 32 instituciones; trayectoria latente del modelo estimado por máxima verosimilitud y potencia por simulación de Monte Carlo. Panel C: potencia estimada por simulación de Monte Carlo (500 réplicas por tamaño; prueba de significación conjunta para los efectos indirectos) en función del tamaño muestral efectivo, con su equivalente en estudiantes (eje superior, efecto de diseño 1.56); la línea discontinua marca el requisito de 250 casos efectivos y la continua el tamaño alcanzado (480/1.56 = 307.7). Potencia del efecto indirecto de BC: .806 con 250 y .854 con 300 casos efectivos.

Por último, la Figura 18-4 resume la relación entre autoeficacia y desempeño por institución, la distribución de las barreras entre las 32 instituciones y las curvas de potencia del diseño.

18.5 DISCUSIÓN

El hallazgo más desconcertante no es la asociación esperada entre prácticas pertinentes y autoeficacia, sino que las barreras contextuales, con un efecto total negativo y distinto de cero sobre el desempeño, no dejaron una huella distinguible en las creencias de eficacia ni en ninguna de sus dos trayectorias directas consideradas por separado, aunque ambas trayectorias directas, tomadas en conjunto, mejoraron el ajuste del modelo. Se sostuvieron cuatro de las seis hipótesis (H1, H3, H4 y H6): las prácticas culturalmente pertinentes se asociaron con mayor autoeficacia y, a través de ella, con mayor desempeño; la autoeficacia se relacionó con el desempeño medido con pruebas; y el modelo resultó equivalente entre estudiantes quechuahablantes y aimarahablantes. H2 y H5 no se sostuvieron, y esa asimetría entre las dos condiciones de enseñanza organiza lo que sigue.

18.5.1 Implicaciones teóricas

Desde una perspectiva crítica, la lectura más prudente de la ausencia de mediación de las barreras es que estas restringen el aprendizaje por vías difusas, como la exposición a la lengua o la continuidad de la instrucción, más que por las creencias de eficacia de los estudiantes. Esta distinción precisa lo que los estudios cualitativos sobre el inglés rural habían descrito como un conjunto indiferenciado de desafíos —docentes generalistas, escasez de materiales, débil articulación con la cultura local— (Izquierdo et al., 2021; Ortiz y Contreras, 2023; Vega et al., 2025): al descomponer los efectos, la carencia de condiciones escolares se asocia con un menor desempeño, pero no parece operar sobre la confianza del estudiante en su propia capacidad. Una explicación complementaria, que los datos sugieren sin permitir contrastarla, es de nivel de análisis. Las barreras son en buena

medida una propiedad de la escuela, con la correlación intraclase más alta de los cuatro constructos, mientras que la autoeficacia varía sobre todo entre estudiantes de una misma institución, con la más baja; un mediador que varía dentro de la escuela difícilmente transmita un efecto que varía entre escuelas, y un modelo de un solo nivel con 32 conglomerados, aun con errores estándar ajustados, no puede separar ambos componentes. A ello se suma que la escala de barreras tuvo una varianza media extraída algo inferior a .50, de modo que cierto atenuamiento por error de medición no puede descartarse. Cabe recordar, por último, que la trayectoria hacia la autoeficacia fue menor que el efecto mínimo de interés fijado a priori, pero su intervalo no excluye ese valor: se trata de ausencia de evidencia suficiente, no de evidencia de ausencia.

En contraste, las prácticas culturalmente pertinentes —lo que el docente puede modificar en el aula— se relacionaron con el desempeño sobre todo a través de la autoeficacia, que transmitió la mitad de su efecto total, y su efecto indirecto superó al de las barreras. Este patrón es consistente con la interpretación social cognitiva de que conectar las tareas con la experiencia del estudiante multiplica las experiencias de dominio y ofrece modelos cercanos, y con la propuesta de que el reconocimiento de la diversidad lingüística y cultural es condición del aprendizaje en la educación bilingüe rural (Duque et al., 2024) y de que la enseñanza del inglés a estudiantes indígenas solo es aditiva si se apoya en la lengua originaria y en el castellano académico (Fabre et al., 2025). Los docentes EIB del norte del Perú que juzgan viable integrar contenidos en lenguas originarias (Montoya et al., 2025) encuentran aquí un correlato desde el lado del estudiante. También es compatible con la evidencia peruana de que las condiciones del lado docente se asocian con el logro de los estudiantes de la escuela pública (Rentería y Solano, 2025). Notablemente, la trayectoria directa de las prácticas, con un intervalo que apenas incluye el cero, sugiere que una fracción de su asociación con el desempeño no pasa por la autoeficacia, quizá porque el uso de la lengua originaria como apoyo facilita la comprensión de las tareas; el diseño no permite adjudicar esa fracción.

La trayectoria entre autoeficacia y desempeño (β = .393) es del orden de la correlación media metaanalítica entre autoeficacia y logro en lenguas adicionales (Goetze y Driver, 2022), aunque aquí se estimó con pruebas de desempeño y no con calificaciones o autoinformes, y se alinea con el papel predictor de la autoeficacia académica documentado en estudiantes universitarios peruanos (Correa et al., 2024) y con su replicación en la escritura (Sun et al., 2021).

La invarianza entre quechuahablantes y aimarahablantes es un resultado en sí mismo. Ambas lenguas difieren en estatus y en presencia escolar (Zavala, 2023; Sullón, 2024), y una expectativa razonable era que la autoeficacia, o su relación con el desempeño, operara de otro modo en la comunidad cuya lengua goza de menor revalorización social; sin embargo, ni los instrumentos ni las trayectorias discriminaron entre grupos. Que los instrumentos midan lo mismo permite comparar medias, y estas mostraron una única diferencia: mayores barreras en el grupo aimara. Esa diferencia exige cautela, porque el estrato aimara procede solo de Puno y su carga horaria de inglés fue menor, de modo que lengua, región y tiempo de instrucción se confunden. Equivalencia, además, no equivale a igualdad exacta: con 240 estudiantes por grupo, el diseño carece de potencia para detectar diferencias pequeñas, y la mayor discrepancia observada, la de la trayectoria directa de las prácticas, no alcanzó la significación. Para un sistema EIB que atiende a varias lenguas originarias (Kvietok y Hornberger, 2023), el resultado sugiere que un mismo modelo es candidato a la generalización, pero conviene ponerlo a prueba con lenguas de menor presencia escolar antes de afirmarlo.

18.5.2 Implicaciones prácticas

Para la política de educación intercultural bilingüe, los resultados sugieren dos líneas complementarias. La primera es pedagógica y está al alcance de cada docente: incorporar la lengua originaria y los saberes de la comunidad en la clase de inglés se asocia con más autoeficacia y, a través de ella, con mejor desempeño. La segunda es institucional: dado que el efecto de las barreras no parece canalizarse por las creencias de

eficacia, es poco probable que las estrategias motivacionales lo corrijan, y las señales apuntan a materiales, continuidad docente y tiempo efectivo de instrucción, limitaciones que reconoce la reforma de la formación docente intercultural bilingüe (Moreano et al., 2026) y que se inscriben en una EIB implantada sobre todo en contextos rurales (Flores et al., 2023). Que las barreras fueran mayores en las escuelas aimaras señala, con la cautela anotada, dónde priorizar la inversión. En cuanto a la formación, la inicial y la de servicio de los docentes de inglés podrían incorporar de forma explícita la enseñanza culturalmente pertinente; la evaluación a escala de un programa de acompañamiento pedagógico en 3797 escuelas rurales peruanas con alta rotación docente (Castro et al., 2025) apunta a que mejorar las prácticas de enseñanza en este entorno es viable.

18.5.3 Limitaciones y líneas futuras

El diseño transversal impide establecer el orden temporal entre condiciones de enseñanza, autoeficacia y desempeño; un diseño longitudinal con mediciones repetidas permitiría contrastar trayectorias en ambas direcciones. Las prácticas y las barreras se midieron con la percepción del estudiante, que puede diferir de la observación del aula; combinarla con registros de observación o con la perspectiva docente fortalecería la inferencia. La escala de barreras contextuales tuvo una varianza media extraída ligeramente inferior a .50, por lo que su revisión, en particular la del indicador de rotación docente, es una tarea pendiente. El tamaño muestral garantizó una potencia de .80 para efectos de la magnitud fijada a priori; los efectos de las barreras resultaron menores y sus intervalos no excluyen ni el cero ni ese valor, de modo que no permiten descartar un efecto pequeño. Por la misma razón, el estudio no tiene potencia para detectar diferencias pequeñas entre grupos en trayectorias concretas. Un análisis multinivel, con más instituciones por grupo, permitiría separar el componente escolar de las barreras del componente individual de la autoeficacia. Por último, la muestra procede de la sierra sur y de dos lenguas originarias; extender el modelo a la Amazonía, con mayor diversidad lingüística, pondría a prueba su generalidad.

18.6 CONCLUSIONES

En la escuela intercultural bilingüe rural andina del Perú, las prácticas didácticas culturalmente pertinentes se relacionan con un mejor desempeño en inglés, y la autoeficacia transmite la mitad de ese efecto: la creencia de eficacia es el mecanismo mejor sustentado por el que la pertinencia cultural de la didáctica se vincula con el aprendizaje de una tercera lengua. Las barreras contextuales se asocian con un menor desempeño, pero no por una vía distinguible a través de la autoeficacia, lo que desplaza la respuesta hacia las condiciones materiales y organizativas de la escuela y no hacia intervenciones centradas en la motivación del estudiante.

El modelo de medición y las trayectorias estructurales son equivalentes entre estudiantes quechuahablantes y aimarahablantes, lo que permite aplicar una misma estrategia pedagógica a ambas comunidades y priorizar la reducción de barreras donde estas son mayores. Para la enseñanza de una lengua internacional como tercera lengua, el estudio aporta evidencia de que la pertinencia cultural de la didáctica no es un complemento identitario, sino una vía con efecto medible sobre el aprendizaje, y ofrece un modelo de medición y una prueba de invarianza que otros sistemas con lenguas originarias pueden replicar.

Las limitaciones señaladas —diseño transversal, medición basada en la percepción del estudiante, una escala de barreras mejorable y una muestra de dos lenguas de la sierra sur— delimitan el alcance de estas conclusiones y definen la agenda inmediata: estudios longitudinales que contrasten el orden temporal, mediciones de aula independientes del autoinforme, diseños multinivel que separen el componente escolar de las barreras y réplicas en la Amazonía, donde la diversidad lingüística es mayor.

18.7 REFERENCIAS BIBLIOGRÁFICAS

Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2023). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics, 138*(1), 1-35. https://doi.org/10.1093/qje/qjac038

Castro, J., Glewwe, P., Heredia, A., Majerowicz, S., & Montero, R. (2025). Can teaching be taught? Improving teachers' pedagogical skills at scale in rural Peru. *Quantitative Economics, 16*(1), 185-233. https://doi.org/10.3982/qe2079

Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cheung, G., Cooper, H., Lau, R., & Wang, L. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management, 41*(2), 745-783. https://doi.org/10.1007/s10490-023-09871-y

Correa, J., Grimaldo, M., Marcelo, E., Manzanares, E., & Ravelo, E. (2024). Growth mindset, Grit, and Academic Self-efficacy as determinants of Academic Performance in Peruvian university students. *International Journal of Educational Psychology*, 1-19. https://doi.org/10.17583/ijep.13115

Curo, M., Sairitupa, L., Peralta, G., Morales, M., Ilyés, R., & Morales, W. (2025). Adaptation and Validation of a Spanish Writing Self-Efficacy Scale in Quechua-Speaking Peruvian Basic Education

Students. *Behavioral Sciences, 15*(10), 1418. https://doi.org/10.3390/bs15101418

Donnelly, S., Jorgensen, T., & Rudolph, C. (2023). Power analysis for conditional indirect effects: A tutorial for conducting Monte Carlo simulations with categorical exogenous variables. *Behavior Research Methods, 55*(7), 3892-3909. https://doi.org/10.3758/s13428-022-01996-0

Duque, D., Tangarife, M., & Velásquez, Á. (2024). Interculturality in Latin American Rural Bilingual Education: A Systematic Literature Review. *Profile: Issues in Teachers' Professional Development, 26*(2), 199-215. https://doi.org/10.15446/profile.v26n2.109822

Elizondo, J., González, A., & Araya, W. (2026). Assessing English language proficiency in Costa Rican high schools. *Studies in Language Assessment, 15*(1), 30-67. https://doi.org/10.58379/lisi3879

Fabre, P., Faytong, M., Contreras, A., Angulo, A., Paez, A., & Fabre, P. (2025). Beyond bilingualism: an additive trilingual equity agenda for Indigenous English education in Latin America. *Frontiers in Education, 10.* https://doi.org/10.3389/feduc.2025.1643377

Fabre, P., Paez, A., Contreras, A., Fabre, P., Angulo, A., & Faytong, M. (2026). The soft power of English: language promotion and public policy in Latin America. *Frontiers in Communication, 11.* https://doi.org/10.3389/fcomm.2026.1643403

Flores, H., Gamarra, J., Mamani, C., & Vilcarromero, L. (2023). Intercultural Bilingual Education (EIB) in Peru. *Research in Subject-*

matter Teaching and Learning (RISTAL), 6(1), 84-95. https://doi.org/10.2478/ristal-2023-0006

Goetze, J., & Driver, M. (2022). Is learning really just believing? A meta-analysis of self-efficacy and achievement in SLA. *Studies in Second Language Learning and Teaching*, *12*(2), 233-259. https://doi.org/10.14746/ssllt.2022.12.2.4

Guerra, M., Reynosa, E., Durand, E., Florez, A., Chambi, M., Vargas, E., & Cari, E. (2024). Enjoyment and oral English proficiency in future teachers. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1306080

Guevara, R., Tongo, D., Rojas, E., & Yaun, M. (2026). Principals' Technological Leadership and School Digital Equity in Rural Indigenous Schools of the Peruvian Amazon. *Leadership and Policy in Schools*, 1-19. https://doi.org/10.1080/15700763.2026.2669722

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huh, I., & Gim, J. (2025). Exploration of Likert scale in terms of continuous variable with parametric statistical methods. *BMC Medical Research Methodology*, *25*(1). https://doi.org/10.1186/s12874-025-02668-1

Izquierdo, J., Aquino, S., & García, V. (2021). Foreign language education in rural schools: Struggles and initiatives among generalist teachers teaching English in Mexico. *Studies in Second Language Learning and Teaching*, *11*(1), 133-156. https://doi.org/10.14746/ssllt.2021.11.1.6

Kalkbrenner, M. (2024). Choosing Between Cronbach's Coefficient Alpha, McDonald's Coefficient Omega, and Coefficient H: Confidence Intervals and the Advantages and Drawbacks of Interpretive Guidelines. *Measurement and Evaluation in Counseling and Development, 57*(2), 93-105. https://doi.org/10.1080/07481756.2023.2283637

Kvietok, F., & Hornberger, N. (2023). Bringing the language forward: engagements with Quechua language planning and policy. *International Journal of the Sociology of Language, 2023*(280), 1-12. https://doi.org/10.1515/ijsl-2022-0017

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & Van De Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research, 110,* 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Mendoza, M., Morales, W., Sairitupa, L., Morales, S., Rivera, O., Sucapuca, F., & Cunza, D. (2023). Psychometric properties and invariance of an English self-efficacy scale for university students in Peru. *Frontiers in Psychology, 14.* https://doi.org/10.3389/fpsyg.2023.1187342

Montoya, G., Díaz, F., & Usquiano, M. (2025). Beyond Educational Policies: Intercultural Bilingual Education Teachers and the Use of Multimedia Content in Indigenous Languages. *Journal of Language Teaching and Research, 16*(5), 1445-1455. https://doi.org/10.17507/jltr.1605.01

Moreano, G., Zegarra, C., & Rivera, Y. (2026). New intercultural bilingual education teacher preparation curriculum: Opportunities and challenges to reduce educational inequities in Peru. *Education Policy Analysis Archives, 34*. https://doi.org/10.14507/epaa.34.9008

Moshagen, M., & Bader, M. (2024). semPower: General power analysis for structural equation models. *Behavior Research Methods, 56*(4), 2901-2922. https://doi.org/10.3758/s13428-023-02254-7

Ortiz, W., & Contreras, A. (2023). Teaching English in the Colombian Rural: Normalistas Superiores Challenges and Experiences. *Enletawa Journal, 16*(2). https://doi.org/10.19053/2011835x.16352

Preacher, K., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. https://doi.org/10.3758/BRM.40.3.879

Preacher, K., & Selig, J. (2012). Advantages of Monte Carlo Confidence Intervals for Indirect Effects. *Communication Methods and Measures, 6*(2), 77-98. https://doi.org/10.1080/19312458.2012.679848

Quispe, R., Arteta, G., Quispe, D., & Morales, W. (2024). Validation of a Spanish version of the Foreign Language Classroom Anxiety Scale

in Peruvian secondary education students. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1394157

Rentería, J., & Solano, D. (2025). The parabolic path of teacher well-being and student learning achievement in Peru. *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-05736-5

Rojas, R., Paz, R., Arpi, R., Quispe, C., & Chura, E. (2024). Urban-rural gap in education performance in Peruvian public institutions during 2018: an analysis using the Oaxaca-Blinder decomposition. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1394938

Sullón, K. (2024). El paisaje lingüístico en las escuelas de Educación Intercultural Bilingüe del Perú. *Revista Iberoamericana de Educación, 96*(1), 83-95. https://doi.org/10.35362/rie9616436

Sun, T., Wang, C., Lambert, R., & Liu, L. (2021). Relationship between second language English writing self-efficacy and achievement: A meta-regression analysis. *Journal of Second Language Writing, 53*, 100817. https://doi.org/10.1016/j.jslw.2021.100817

Tarrillo, J., Mamani, L., Tocto, E., Iftikhar, H., Gonzales, R., & Soria, J. (2026). Exploring patterns in intercultural bilingual education in Peru using machine learning. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1832249

Vega, J., Correa, A., Marguart, C., & Vargas, M. (2025). Pre-service teachers' perceptions about English language teaching in rural schools: A qualitative study in a Chilean context. *Logos*. https://doi.org/10.15443/rl3536

Zavala, V. (2023). Youth, Quechua and neoliberalism in contemporary Perú. *International Journal of the Sociology of Language*, *2023*(280), 45-66. https://doi.org/10.1515/ijsl-2022-0020


NOVO MUNDO