Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), septiembre-octubre 2026,
Volumen 10, Número 5.

https://doi.org/10.37811/cl_rcm.v10i5

EFECTO DE UN PROGRAMA DE DELIBERACIÓN PEDAGÓGICA ESTRUCTURADA SOBRE EL PENSAMIENTO CRÍTICO EN LA FORMACIÓN INICIAL DOCENTE: ESTUDIO CUASIEXPERIMENTAL

EFECTO DE UN PROGRAMA DE DELIBERACIÓN PEDAGÓGICA ESTRUCTURADA SOBRE EL PENSAMIENTO CRÍTICO EN LA FORMACIÓN INICIAL DOCENTE: ESTUDIO CUASIEXPERIMENTAL

Vilma Lilian Alvarado Orbegoso
Universidad Privada Antenor Orrego, Perú

Karina Lizet Salinas Carranza
Universidad Privada Antenor Orrego, Perú

DOI: https://doi.org/10.37811/cl_rcm.v10i5.26742

Efecto de un Programa de Deliberación Pedagógica Estructurada sobre el Pensamiento Crítico en la Formación Inicial Docente: Estudio Cuasiexperimental

Vilma Lilian Alvarado Orbegoso [1]
valvaradoo@upao.edu.pe
https://orcid.org/0009-0003-1846-1120
Universidad Privada Antenor Orrego
Trujillo, Perú

Karina Lizet Salinas Carranza
ksalinasc@upao.edu.pe
https://orcid.org/0000-0002-1835-3950
Universidad Privada Antenor Orrego
Trujillo, Perú

RESUMEN

El desarrollo del pensamiento crítico en la formación docente exige experiencias en las que las futuras profesoras no solo conozcan contenidos, sino que deban examinar evidencias, contrastar razones y justificar decisiones pedagógicas. El objetivo de este estudio fue determinar el efecto de un programa de deliberación pedagógica estructurada sobre el pensamiento crítico de estudiantes del IV ciclo de Educación Inicial de una escuela pedagógica pública de Trujillo, Perú. Se realizó una investigación cuantitativa, aplicada y de nivel explicativo, con diseño cuasiexperimental de pretest-postest y grupo control no equivalente. Participaron 56 estudiantes mediante censo, 28 en el grupo experimental y 28 en el control. El desenlace se evaluó con la Prueba de Pensamiento Crítico en Situaciones Pedagógicas (PPC-SP), compuesta por 20 ítems contextualizados en cuatro casos de Educación Inicial. Antes de la intervención no se halló diferencia significativa entre los grupos ($U = 310.00$; $p = .173$). Después de diez sesiones, el grupo experimental alcanzó una media de 17.75 puntos y el control 9.36 ($U = 784.00$; $p < .001$). El ANCOVA, con el pretest como covariable, estimó una diferencia ajustada de 8.52 puntos, $F(1, 53) = 1151.13$, $p < .001$, $\eta p^2 = .956$. En las cinco dimensiones se observaron igualmente diferencias significativas. En el contexto de esta intervención, el grupo que participó en la deliberación pedagógica estructurada mostró un aumento considerable del pensamiento crítico. Este resultado debe leerse con prudencia, dado que proviene de un diseño cuasiexperimental con dos conglomerados, utiliza un instrumento situado y presenta efecto techo en algunas dimensiones.

Palabras clave: pensamiento crítico; deliberación pedagógica; formación inicial docente; Educación Inicial; estudio cuasiexperimental

[1] Autor principal.
Correspondencia: valvaradoo@upao.edu.pe

Efecto de un Programa de Deliberación Pedagógica Estructurada sobre el Pensamiento Crítico en la Formación Inicial Docente: Estudio Cuasiexperimental

ABSTRACT

Critical thinking in teacher education is strengthened when future teachers are required not only to learn pedagogical content but also to weigh evidence, compare reasons, and justify professional decisions. This study aimed to determine the effect of a structured pedagogical deliberation program on critical thinking among fourth-semester Early Childhood Education students at a public teacher-training school in Trujillo, Peru. A quantitative, applied, explanatory quasi-experimental pretest-posttest design with a non-equivalent control group was used. The census included 56 students, 28 in the experimental group and 28 in the control group. The outcome was assessed with the Critical Thinking Test in Pedagogical Situations, a 20-item instrument organized around four Early Childhood Education cases. No significant baseline difference was found between groups ($U = 310.00$; $p = .173$). After ten sessions, the experimental group reached a mean score of 17.75, compared with 9.36 in the control group ($U = 784.00$; $p < .001$). ANCOVA using pretest scores as a covariate yielded an adjusted difference of 8.52 points, $F(1, 53) = 1151.13$, $p < .001$, partial $\eta^2 = .956$. Significant posttest differences were also observed across the five critical-thinking dimensions. Under the conditions examined, structured pedagogical deliberation was associated with a substantial improvement in critical thinking. The magnitude of the effect should nevertheless be interpreted cautiously because of the quasi-experimental design, the limited number of clusters, the context-specific assessment, and ceiling effects in some dimensions.

Keywords: critical thinking; pedagogical deliberation; initial teacher education; early childhood education; quasi-experimental study

Artículo recibido 20 agosto 2026
Aceptado para publicación: 20 setiembre 2026

INTRODUCCIÓN

La formación inicial docente enfrenta un problema que no se resuelve únicamente ampliando el volumen de contenidos: las futuras profesoras necesitan aprender a decidir cuándo la situación educativa admite más de una respuesta plausible. En el ejercicio profesional, esto supone separar hechos de interpretaciones, ponderar la solidez de los argumentos, anticipar posibles consecuencias y estar dispuesta a revisar una decisión cuando aparece nueva evidencia. En Educación Inicial, este razonamiento cobra especial importancia porque muchas decisiones vinculadas con convivencia, inclusión, diversidad cultural y desarrollo infantil no se resuelven mediante la aplicación automática de una norma. El Diseño Curricular Básico Nacional peruano propone desarrollar competencias relacionadas con el análisis crítico, la convivencia democrática y la deliberación (Ministerio de Educación del Perú [MINEDU], 2019). Sin embargo, no basta con que estos propósitos figuren en el documento curricular; en la formación se necesitan situaciones concretas en las que las estudiantes practiquen este razonamiento con frecuencia y lo usen al tomar decisiones pedagógicas.

Para analizar este desempeño, se tomó como referencia el concepto de pensamiento crítico de Facione (1990), quien lo definió como un juicio intencional y autorregulado que comprende seis habilidades: interpretación, análisis, evaluación, inferencia, explicación y autorregulación. Este concepto es útil para la investigación porque no se queda en una actitud general, sino que precisa los procesos cognitivos que una persona utiliza cuando elabora o revisa una postura. Estudios posteriores mostraron que estas habilidades no aparecen de manera espontánea; su desarrollo depende de la forma en que se vinculan con los contenidos de cada disciplina y con las tareas que se proponen en clase. El metaanálisis de Abrami et al. (2015) encontró un efecto global positivo y resaltó el uso de estrategias explícitas, el diálogo y la integración de estas habilidades con los contenidos disciplinares. Por esta razón, hoy la discusión ya no se centra en si el pensamiento crítico puede enseñarse, sino en qué condiciones permiten que las estudiantes lo apliquen ante problemas reales.

Una de estas condiciones es el diálogo basado en razones. Ten Dam y Volman (2004) vincularon el pensamiento crítico con el uso de metodologías activas, el trabajo con problemas reales y la interacción entre pares. Alexander (2020), desde la enseñanza dialógica, señala que la conversación en el aula tiene valor formativo cuando es colectiva, recíproca, acumulativa y tiene un propósito claro.

Por otro lado, Reznitskaya y Gregory (2013) encontraron que el razonamiento de los estudiantes cambia cuando el docente deja de concentrar la palabra y les pide responder a los argumentos de sus compañeros. Esta idea se relaciona con la comunidad de indagación de Lipman (2003) y con la ética discursiva de Habermas (1981/1999). En ambos casos, deliberar no significa sumar opiniones, sino exponer las propias ideas ante los demás para que sean examinadas con preguntas, razones y objeciones, y revisarlas si es necesario.

En educación superior también hay estudios que apoyan esta línea. Zhang et al. (2022) hallaron mejoras importantes cuando la *enseñanza* del pensamiento crítico se vinculó con el contenido disciplinar. Cai et al. (2022) obtuvieron resultados significativos con secuencias que incluían estudio individual, discusión y trabajo colaborativo, y Wu et al. (2023) reportaron ventajas al combinar observación, instrucción directa, debate y microenseñanza. Orhan (2024), Bates et al. (2025) y Song et al. (2025), en distintos países, encontraron que el trabajo con problemas, la evaluación de evidencia, la escritura argumentativa y el diálogo reflexivo favorecen el razonamiento crítico. En formación docente, Ossa Cornejo et al. (2025) registraron avances significativos luego de aplicar talleres específicos. En el Perú se han reportado resultados parecidos con aprendizaje situado, programas estructurados y estrategias de discusión (Alcarraz Carbajal, 2024; Padilla Sánchez et al., 2024; Valencia Morocho, 2021).

A pesar de estos avances, todavía hay aspectos poco estudiados en la formación inicial docente. Muchos trabajos utilizan pruebas estandarizadas generales, que no permiten ver cómo razonan las estudiantes cuando deben tomar decisiones *pedagógicas* concretas. Además, en las escuelas de educación superior pedagógica del Perú hay pocos estudios cuasiexperimentales, y se sabe poco sobre programas deliberativos que formen parte del currículo regular y que trabajen a la vez el examen de evidencia, la respuesta a objeciones y la autorregulación. En la institución donde se realizó este estudio, el diagnóstico formativo mostró que las estudiantes tenían dificultades para distinguir datos de interpretaciones, cuestionar argumentos sostenidos solo en la costumbre o la autoridad y justificar sus decisiones pedagógicas sin limitarse a citar la norma. Es decir, el problema no era la falta de contenidos, sino las pocas oportunidades para razonar sobre ellos y discutirlos con otras personas.

Para atender esta necesidad, se diseñó y aplicó un programa de deliberación pedagógica estructurada en el curso *Deliberación y Participación*.

El programa se organizó a partir del análisis de dilemas reales de Educación Inicial, con pautas para el intercambio de ideas, una secuencia de preguntas guía y un cierre de reflexión metacognitiva. Se buscó que el razonamiento dejara de ser una actividad ocasional y se convirtiera en una práctica constante, en la que las estudiantes identificaran lo que se sabe, anticiparan consecuencias, compararan argumentos, justificaran sus decisiones y reconocieran qué evidencia las haría cambiar de postura. Por ello, el objetivo general fue evaluar el efecto del programa sobre el pensamiento crítico global y sus dimensiones de análisis, inferencia, evaluación, explicación y autorregulación. La hipótesis planteó que las estudiantes del grupo experimental obtendrían en el postest un desempeño significativamente mayor que las que trabajaron los mismos contenidos con la metodología tradicional.

METODOLOGÍA

Diseño y participantes

El estudio tuvo enfoque cuantitativo, fue de tipo aplicado y de nivel explicativo, con diseño cuasiexperimental de pretest y postest con grupo control no equivalente (Campbell & Stanley, 1963; Hernández-Sampieri & Mendoza, 2018). Las estudiantes ya estaban distribuidas en dos secciones al empezar el semestre, por lo que no fue posible asignarlas al azar de manera individual. Para evitar que la investigadora eligiera qué sección recibiría el programa, la condición experimental se asignó por sorteo. Este procedimiento no convierte el estudio en un experimento puro, pero reduce la intervención de la investigadora en la asignación y mantiene los grupos tal como estaban conformados.

La población estuvo formada por las 56 estudiantes matriculadas en el IV ciclo de Educación Inicial, en el curso *Deliberación y Participación*, durante el semestre 2026-I. Todas eran mujeres y tenían entre 18 y 25 años. Se trabajó con toda la población: 28 estudiantes integraron el grupo experimental y 28 el grupo control. Se incluyó a quienes tenían matrícula regular, firmaron el consentimiento informado y resolvieron completo el pretest, y se excluyó a las estudiantes repitentes, con licencia académica o con matrícula condicionada. También se definió que se eliminarían los registros de quienes faltaran a más del 30 % de las sesiones, no rindieran el postest o dejaran sin responder o con doble marca más del 10 % de los ítems. No se presentó ninguno de estos casos, por lo que los 56 registros entraron al análisis. Por tanto, los resultados solo pueden extenderse a esta población y a este contexto institucional.

Intervención

El programa se desarrolló en diez sesiones semanales de 90 minutos, entre la segunda y la undécima semana del semestre. Se aplicó dentro del horario regular del curso y ambos grupos trabajaron los mismos contenidos del sílabo, con los mismos productos evaluables y la misma carga horaria; la diferencia estuvo en la metodología. En el grupo *experimental*, cada sesión tuvo cinco momentos: encuadre y activación de saberes previos, presentación de un dilema pedagógico, deliberación en pequeños grupos, plenario y cierre metacognitivo.

En todas las sesiones del grupo experimental se utilizó la estrategia *"Ruta de las cinco preguntas"*, elaborada en función de las dimensiones de pensamiento crítico que se iban a evaluar. Frente a cada dilema, las estudiantes debían separar los hechos de los supuestos, prever consecuencias, valorar la consistencia de los argumentos, justificar su postura e indicar qué evidencia las llevaría a cambiar su decisión. El grupo control trabajó los mismos contenidos con las estrategias que se usan habitualmente en el curso. Un observador externo registró la fidelidad de la implementación con una lista de cotejo de nueve criterios.

Instrumento

Para medir la variable dependiente, se construyó la *Prueba de Pensamiento Crítico en Situaciones Pedagógicas* (PPC-SP), un instrumento de desempeño elaborado *ad hoc* para esta investigación. Tiene 20 ítems de opción múltiple con cuatro alternativas, organizados en cuatro casos ambientados en aulas de Educación Inicial. Para que ninguna dimensión dependiera de un solo caso, la tabla de especificaciones se armó de forma cruzada: cada caso incluye un ítem de análisis, uno de inferencia, uno de *evaluación*, uno de explicación y uno de autorregulación.

El puntaje total va de 0 a 20 puntos y cada dimensión se califica de 0 a 4. Para disminuir la deseabilidad social y evitar que la respuesta correcta fuera evidente, los distractores se redactaron a partir de errores de razonamiento frecuentes en la formación docente, como la generalización apresurada, el uso de la experiencia personal como única prueba, el argumento circular y la aceptación de lo que opina la mayoría.

La validez de contenido fue evaluada por cinco jueces con grado de doctor; tres eran especialistas en Educación Inicial y didáctica de la formación docente, y dos en metodología y psicometría. Cada juez

calificó la claridad, coherencia, relevancia y suficiencia de los ítems. Se conservaron los que obtuvieron una V de Aiken ≥ .80 y un límite inferior del intervalo de confianza al 95 % ≥ .59 (Aiken, 1985; Escurra, 1988). Luego se aplicó una prueba piloto a 30 estudiantes con características equivalentes, ajenas a la población del estudio. En esta fase se examinaron la dificultad, la discriminación y el funcionamiento de los distractores; se conservaron los reactivos con índices de dificultad entre .30 y .85 y correlación biserial puntual corregida ≥ .20. La consistencia del puntaje global se estimó con KR-20 (Kuder & Richardson, 1937). Debido a que cada subescala estuvo integrada por cuatro ítems dicotómicos, los análisis por dimensión se asumieron como exploratorios.

Procedimiento, análisis y consideraciones éticas

Durante la primera semana, una persona externa capacitada administró el pretest a ambos grupos de manera simultánea, utilizando una consigna estandarizada, hojas codificadas y un tiempo de 40 minutos. La intervención se desarrolló durante diez semanas y el postest se administró en la semana 12 bajo las mismas condiciones. La investigadora fue, simultáneamente, docente del curso y autora del programa; para reducir el riesgo asociado a este doble rol, la administración de las pruebas se delegó, la corrección se realizó de forma ciega respecto del grupo y del momento de medición y un observador independiente verificó la implementación. Los resultados de la PPC-SP no se utilizaron para la calificación académica. El procesamiento se efectuó en IBM SPSS Statistics 27 y se verificó con Jamovi 2.3. Primero se obtuvieron los estadísticos descriptivos del puntaje global y de cada dimensión. Para la normalidad se usó la prueba de Shapiro-Wilk y para la homogeneidad de varianzas, la de Levene. Cuando no se cumplieron los supuestos, las diferencias entre grupos se analizaron con la U de Mann-Whitney y los cambios dentro de cada grupo con la prueba de rangos con signo de Wilcoxon. En las comparaciones por dimensión se aplicó la corrección de Bonferroni ($\alpha = .01$). El análisis principal fue un ANCOVA con postest como variable dependiente, grupo como factor y pretest como covariable, procedimiento apropiado para reducir parcialmente el efecto de diferencias iniciales en diseños con grupos no equivalentes (Shadish et al., 2002). Se reportaron g de Hedges, r y eta cuadrado parcial como medidas de magnitud, con $\alpha = .05$. La interpretación del ANCOVA se realizó conjuntamente con el contraste no paramétrico debido a las desviaciones de normalidad observadas.

El estudio contó con aprobación del Comité de Ética de la Escuela de Posgrado de la Universidad Privada Antenor Orrego. Las participantes firmaron el consentimiento informado y podían retirarse del estudio sin que esto afectara sus notas. Las respuestas se identificaron mediante códigos, la correspondencia entre códigos e identidades se mantuvo bajo custodia restringida y los datos se utilizaron únicamente con fines académicos. Los casos de la intervención y del instrumento no reprodujeron situaciones identificables de niñas, familias o instituciones reales. Una vez concluido el postest, el grupo control recibió un taller de retroalimentación con los componentes centrales del programa.

RESULTADOS Y DISCUSIÓN

Cambio global del pensamiento crítico

Tabla 1 Puntaje global y niveles de pensamiento crítico según grupo y momento

Grupo y momento	M	DE	Inicio n (%)	Proceso n (%)	Logro esperado n (%)	Logro destacado n (%)
Experimental, pretest	7.11	2.01	23 (82.1)	5 (17.9)	0 (0.0)	0 (0.0)
Control, pretest	7.39	1.07	28 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Experimental, postest	17.75	1.21	0 (0.0)	0 (0.0)	15 (53.6)	13 (46.4)
Control, postest	9.36	1.16	19 (67.9)	9 (32.1)	0 (0.0)	0 (0.0)

Nota. M = media; DE = desviación estándar. N = 28 por grupo.

En el pretest, ambos grupos mostraron un desempeño bajo. En el grupo experimental, 82.1 % de las estudiantes se ubicó en Inicio, mientras que en el control lo hizo el 100.0 %; ninguna estudiante alcanzó Logro esperado ni Logro destacado. Los puntajes globales fueron próximos (M = 7.11; DE = 2.01 en el experimental y M = 7.39; DE = 1.07 en el control) y la diferencia no fue estadísticamente significativa (U = 310.00; p = .173; g = -0.18; IC 95 % [-0.70, 0.35]). Este resultado no permite afirmar que las dos secciones fueran equivalentes en todos sus rasgos, aunque sí muestra que el grupo asignado después a la intervención no presentaba una ventaja global detectable al inicio. Este punto de partida coincide con otros estudios recientes con estudiantes de formación docente en el Perú, que también encontraron un amplio margen de mejora antes de la intervención (Sánchez Heredia et al., 2023; Padilla Sánchez et al., 2024).

Después de las diez sesiones, los resultados de ambos grupos fueron muy diferentes. En el grupo experimental, el 53.6 % llegó a Logro esperado y el 46.4 % a Logro destacado; en el control, el 67.9 % siguió en Inicio y el 32.1 % pasó a Proceso. El grupo experimental ganó en promedio 10.64 puntos (DE=1.34), mientras que el grupo control subió 1.96 puntos (DE=1.20). Como las dos secciones desarrollaron los mismos contenidos, esta diferencia apoya la hipótesis de que la metodología influyó en el resultado. Probablemente contribuyeron el análisis frecuente de dilemas, la necesidad de explicar el propio razonamiento, la discusión de objeciones entre compañeras y el momento de reflexión metacognitiva al final de cada sesión. Esto coincide con estudios previos que señalan que el pensamiento crítico mejora cuando se trabaja junto con los contenidos de la disciplina, a partir de problemas reales y con discusiones organizadas (Cai et al., 2022; Orhan, 2024; Ossa Cornejo et al., 2025; Zhang et al., 2022).

La prueba de Wilcoxon mostró que el puntaje global aumentó de forma significativa entre el pretest y el postest en el grupo experimental (W=0.00; p<.001) y también en el grupo control (W=0.00; p<.001). Es decir, ambos grupos mejoraron durante el semestre, pero el aumento fue mucho mayor en la sección que participó en el programa. Aun así, que un grupo mejore respecto de sí mismo no basta para atribuir el cambio al programa; este resultado debe leerse junto con las comparaciones entre grupos y tomando en cuenta que se trabajó con una sola sección por condición.

Efectos por dimensión

Tabla 2 Desempeño por dimensión y comparación postest entre grupos

Dimensión	GE pre M (DE)	GE post M (DE)	GC pre M (DE)	GC post M (DE)	U post	p ajustada	r
Análisis	2.64 (0.78)	4.00 (0.00)	2.57 (0.96)	2.82 (0.67)	728.00	< .001	.83
Inferencia	1.43 (0.74)	3.89 (0.31)	1.61 (0.79)	1.96 (0.64)	776.50	< .001	.90
Evaluación	0.93 (0.81)	3.25 (0.44)	1.07 (0.77)	1.61 (0.63)	763.00	< .001	.85
Explicación	1.54 (1.10)	3.46 (0.51)	1.50 (1.04)	1.96 (1.04)	695.00	< .001	.70
Autorregulación	0.57 (0.63)	3.14 (0.45)	0.64 (0.73)	1.00 (0.72)	780.50	< .001	.89

Nota. GE = grupo experimental; GC = grupo control. La corrección de Bonferroni estableció α = .01 para cada contraste. Como cada subescala tiene solo cuatro ítems, los resultados por dimensión son exploratorios.

En el postest, el grupo experimental obtuvo mejores resultados en las cinco dimensiones, y las diferencias siguieron siendo significativas después de aplicar la corrección de Bonferroni. Los tamaños del efecto (r) fueron altos, entre .70 y .90. Que todas las dimensiones mejoraran a la vez puede explicarse por la forma en que se organizó el programa: en cada dilema, las estudiantes tenían que reconocer la evidencia disponible, inferir consecuencias, valorar los argumentos, justificar una decisión pedagógica y revisar sus propios sesgos. Las habilidades no se trabajaron por separado, sino dentro de situaciones en las que debían decidir como futuras docentes. Esto concuerda con Facione (1990, 2023), quien entiende estas habilidades como partes de un mismo juicio intencional, y con Lipman (2003), para quien el razonamiento se fortalece en comunidades de indagación donde las ideas se presentan, se discuten y se reformulan frente a otros.

Dos dimensiones merecen un comentario aparte. La dimensión *Autorregulación* tuvo una de las medias más bajas en el pretest ($M = 0.57$) y alcanzó 3.14 puntos en el postest del grupo experimental. Este aumento es coherente con la importancia que los estudios dan al monitoreo del propio pensamiento (Ku & Ho, 2010; Rivas et al., 2022). Hay que considerar que todas las sesiones terminaban con un momento metacognitivo en el que cada estudiante revisaba qué había cambiado en su juicio y qué evidencia la llevaría a modificar su postura. La dimensión *Explicación* también aumentó de forma clara. En el programa, las estudiantes tenían que exponer y defender su decisión ante el grupo, lo que se acerca a la ética discursiva de Habermas (1981/1999) y a la enseñanza dialógica de Alexander (2020), donde los argumentos se presentan para que otros los examinen y cuestionen. Bates et al. (2025) encontraron resultados en la misma línea con escritura argumentativa basada en evidencia, debate formativo y diálogo socrático.

Sin embargo, los resultados por dimensión deben leerse con cuidado. En la subescala *Análisis*, la totalidad de las 28 participantes del grupo experimental alcanzó el puntaje máximo ($M = 4.00$; $DE = 0.00$), mientras que en *Inferencia* la media rozó el límite superior ($M = 3.89$; $DE = 0.31$). Este efecto techo hace que el instrumento no pueda diferenciar bien los niveles más altos de desempeño y puede afectar el cálculo del tamaño del efecto. Además, cada dimensión se midió con solo cuatro ítems, por lo que estos resultados complementan el análisis del puntaje global, pero no demuestran que las habilidades se transfieran a otros contextos.

En próximos estudios convendría incluir ítems más difíciles y otras mediciones externas para comprobar si el razonamiento deliberativo se mantiene y se aplica en situaciones nuevas.

Efecto global ajustado y robustez del hallazgo

Tabla 3 Contrastes principales del puntaje global de pensamiento crítico

Contraste	Estadístico	p	Magnitud	Estimación complementaria
GE vs. GC, pretest	U = 310.00	.173	g = -0.18	IC 95 % [-0.70, 0.35]
GE vs. GC, postest	U = 784.00	< .001	g = 6.99	IC 95 % [5.57, 8.41]
ANCOVA: efecto de grupo	F(1, 53) = 1151.13	< .001	ηp² = .956	Diferencia ajustada = 8.52; IC 95 % [8.03, 9.02]

Nota. El ANCOVA tomó el postest como variable dependiente y el pretest como covariable. Medias ajustadas: GE = 17.82; GC = 9.29.

Según la prueba de Shapiro-Wilk, los puntajes globales no siguieron una distribución normal, y la prueba de Levene mostró que las varianzas no eran homogéneas en el pretest ($F = 6.50$; $p = .014$). Por ello, las comparaciones entre grupos se hicieron primero con la prueba no paramétrica U de Mann-Whitney. Aun así, se mantuvo el análisis de covarianza (ANCOVA) como prueba principal porque permite controlar las diferencias iniciales entre las participantes. Se revisaron sus supuestos y se comprobó la homogeneidad de las pendientes de regresión (interacción grupo × pretest: $p = .996$), aunque los residuos se alejaron ligeramente de la normalidad (Shapiro-Wilk $p = .038$). Por este motivo, el modelo debe interpretarse con cautela; sin embargo, el ANCOVA y las pruebas no paramétricas llegaron a los mismos resultados, de modo que las conclusiones no dependen de un solo procedimiento. Luego de ajustar los puntajes según el pretest, la media estimada fue de 17.82 puntos en el grupo experimental y de 9.29 en el grupo control. Esta diferencia ajustada entre ambas condiciones fue de 8.52 puntos (IC 95 % [8.03, 9.02]), situándose en un nivel de significación altamente determinante, $F(1,53) = 1151.13$, $p < .001$, con una magnitud del efecto de gran tamaño ($\eta_p^2 = .956$). Por su parte, la prueba no paramétrica aplicada a los puntajes de postest confirmó de forma consonante esta misma tendencia ($U = 784.00$; $p < .001$; $g = 6.99$; IC 95 % [5.57, 8.41]). Este tamaño del efecto es mucho mayor que el promedio informado en estudios anteriores (v. g., Abrami et al., 2015), por lo que no debe tomarse como un resultado esperable en cualquier programa basado en la deliberación. En este estudio, la intensidad del programa, su relación directa con el curso y el uso de un instrumento elaborado con situaciones de Educación Inicial pudieron hacer que el efecto observado sea mayor.

Aun con estas precauciones, la dirección de los resultados coincide con lo reportado en otros estudios. Reznitskaya y Gregory (2013) señalaron que el razonamiento avanza más rápido en aulas donde el docente no controla toda la conversación y pide a los estudiantes que respondan a las ideas de sus compañeros. Song et al. (2025) encontraron ventajas al integrar el pensamiento crítico con los contenidos del curso, y Ossa Cornejo et al. (2025) registraron avances significativos en futuros docentes que participaron en talleres específicos. En el Perú, Alcarraz Carbajal (2024) y Padilla Sánchez et al. (2024) obtuvieron resultados favorables con intervenciones de aprendizaje situado. Lo que aporta este estudio es la organización de estos principios en una secuencia deliberativa basada en dilemas propios del nivel inicial. Los resultados sugieren que la mejora no se debe solo a que las estudiantes conversaron más en clase, sino a que la conversación siguió reglas claras: sustentar las afirmaciones con evidencia, responder a las objeciones, justificar las decisiones y revisar las propias ideas.

Limitaciones e implicancias

Los resultados deben interpretarse dentro de los límites del diseño. Como se trabajó con dos secciones ya formadas y una sola sección por condición, no es posible separar por completo el efecto del programa de otras características propias de cada aula. Además, la investigadora diseñó y aplicó el programa; aunque las pruebas fueron administradas por una persona externa, la corrección fue ciega y un observador independiente revisó las sesiones, no se puede descartar del todo la influencia de este doble rol. También debe considerarse que la PPC-SP se elaboró *ad hoc* con situaciones parecidas a las trabajadas en la intervención. Esto le da mayor validez ecológica y sensibilidad, pero hace más difícil comparar los resultados con los de pruebas estandarizadas generales. Por último, el efecto techo en análisis e inferencia, la medición únicamente al terminar el programa y el hecho de trabajar en una sola institución limitan la generalización de los hallazgos.

Aun con estas limitaciones, los resultados son útiles para la formación docente. El estudio muestra que el pensamiento crítico puede desarrollarse dentro de los cursos regulares, sin necesidad de crear una asignatura aparte, siempre que las actividades trabajen de manera explícita y frecuente las operaciones del razonamiento. Para la formación inicial docente, la experiencia indica que conviene combinar cuatro elementos: dilemas profesionales reales, reglas claras para el diálogo, preguntas que activen las distintas dimensiones del pensamiento crítico y momentos metacognitivos en cada sesión. Antes de ampliar la

propuesta, será necesario replicarla en varias instituciones, con más de una sección por condición, con docentes facilitadores distintos de la investigadora, con instrumentos externos estandarizados y con mediciones de seguimiento que permitan saber si el razonamiento deliberativo se mantiene en el tiempo y se aplica en la práctica docente. También hace falta aumentar la dificultad de los ítems de análisis e inferencia para reducir el efecto techo y diferenciar mejor los niveles altos de desempeño.

CONCLUSIONES

En las condiciones en que se realizó este estudio, el programa de deliberación pedagógica estructurada se asoció con una mejora importante y estadísticamente significativa del pensamiento crítico en estudiantes de Educación Inicial. Al inicio, los dos grupos tenían puntajes globales parecidos y la mayoría se ubicaba en el nivel *Inicio*. Después de las diez sesiones, todas las estudiantes del grupo experimental alcanzaron *Logro esperado* o *Logro destacado*, mientras que la mayoría del grupo control se mantuvo en los niveles más bajos. El ANCOVA, ajustado por el pretest, mostró una diferencia de 8.52 puntos a favor del grupo experimental.

La mejora se observó en las cinco dimensiones: análisis, inferencia, evaluación, explicación y autorregulación. Los resultados del puntaje global, de las subescalas, de las comparaciones intragrupo y del ANCOVA apuntan en la misma dirección, lo que indica que el cambio no se limitó a una sola habilidad, sino que abarcó el juicio crítico en conjunto. Desde el punto de vista pedagógico, no basta con enseñar contenidos sobre convivencia, diversidad o democracia; las futuras docentes necesitan enfrentarse a situaciones en las que tengan que evaluar evidencia, comparar razones, responder a objeciones y revisar sus propios supuestos.

Por último, el tamaño del efecto encontrado debe leerse con prudencia. El diseño cuasiexperimental, el trabajo con solo dos aulas, el uso de una prueba contextualizada y el efecto techo en algunas subescalas no permiten generalizar estos resultados sin reservas. Este estudio ofrece una evidencia inicial, situada en un contexto específico, y no una estimación general de lo que puede lograr la deliberación pedagógica. Se recomienda que futuras investigaciones repliquen la propuesta en otras instituciones formadoras, con ítems que discriminen mejor, mediciones externas y evaluaciones de seguimiento, para saber si estos avances se mantienen en el tiempo y se reflejan en el desempeño profesional en las aulas de Educación Inicial.

REFERENCIAS BIBLIOGRAFICAS

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314. https://doi.org/10.3102/0034654314551063

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurem*ent, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Alcarraz Carbajal, B. (2024). El aprendizaje situado para desarrollar el pensamiento crítico en las estudiantes de Educación Superior Pedagógica. *Revista Educación*, 22(23), 25–37. https://doi.org/10.51440/unsch.revistaeducacion.2024.23.480

Alexander, R. J. (2020). A dialogic teaching companion. Routledge. https://doi.org/10.4324/9781351040143

Bates, J., Cheng, S., Ferris, M., & Wang, X. (2025). Cultivating critical thinking skills: A pedagogical study in a business statistics course. *Journal of Statistics and Data Science Education*, 33(2), 166–176. https://doi.org/10.1080/26939169.2024.2394534

Cai, X., Peng, M., Qin, J., Zhou, K., Li, Z., Yang, S., & Yan, F. (2022). Sandwich teaching improved students' critical thinking, self-learning ability, and course experience in the Community Nursing Course: A quasi-experimental study. *Frontiers in Psychology,* 13, 957652. https://doi.org/10.3389/fpsyg.2022.957652

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Rand McNally.

Escurra, L. M. (1988). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología,* 6(1-2), 103–111. https://doi.org/10.18800/psico.198801-02.008

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). American Philosophical Association. https://eric.ed.gov/?id=ED315423

Facione, P. A. (2023). Critical thinking: What it is and why it counts (2023 update). Insight Assessment. https://insightassessment.com/wp-content/uploads/2023/12/Critical-Thinking-What-It-Is-and-Why-It-Counts.pdf

Habermas, J. (1999). Teoría de la acción comunicativa (M. Jiménez Redondo, Trad.). Taurus. (Obra original publicada en 1981)

Hernández-Sampieri, R., & Mendoza, C. P. (2018). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill Education.

Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. Metacognition and Learning, 5(3), 251–267. https://doi.org/10.1007/s11409-010-9060-6

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160. https://doi.org/10.1007/BF02288391

Lipman, M. (2003). Thinking in education (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840272

Ministerio de Educación del Perú. (2019). Diseño Curricular Básico Nacional de la Formación Inicial Docente: Programa de estudios de Educación Inicial. MINEDU. https://www.minedu.gob.pe/superiorpedagogica/producto/dcbn-2019-educacion-inicial/

Orhan, A. (2024). Online or in-class problem based learning: Which one is more effective in enhancing learning outcomes and critical thinking in higher education EFL classroom? Journal of Computer Assisted Learning, 40(5), 2351–2368. https://doi.org/10.1111/jcal.13033

Ossa Cornejo, C., Rivas, S. F., & Saiz Sánchez, C. (2025). Effects of two training workshops upon university students' learning of critical thinking. PLOS ONE, 20(1), e0316760. https://doi.org/10.1371/journal.pone.0316760

Padilla Sánchez, M. M., Aranda Sanabria, J. F., Londoñe Cossio, L. M., Gutiérrez Gamboa, E., & Vivas Juan de Dios, S. P. (2024). Impacto de un programa de desarrollo del pensamiento crítico en estudiantes universitarios peruanos. *Socialium*, 8(2), 20–31. https://doi.org/10.26490/uncp.sl.2024.8.2.2018

Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, 48(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, 13, 913219. https://doi.org/10.3389/fpsyg.2022.913219

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Song, Q., Liu, Y., & Zhang, Q. (2025). Effects of critical thinking instruction on Chinese college students with varying baseline critical thinking abilities: A quasi-experimental study. British Educational *Research Journal*, 51(3), 1445–1468. https://doi.org/10.1002/berj.4131

Sánchez Heredia, N., Llaque Fernández, G. I., Valderrama Puscan, M. W., Pérez Urrutia, C. L., & Jara Aquise, A. M. (2023). DAR strategy, with a neuroscientific approach, improves the critical thinking of higher education student. LACCEI, 1(8). https://doi.org/10.18687/LACCEI2023.1.1.1100

Ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: Teaching strategies. *Learning and Instruction*, 14(4), 359–379. https://doi.org/10.1016/j.learninstruc.2004.01.005

Valencia Morocho, C. A. (2021). La educación virtual en el pensamiento crítico de los estudiantes universitarios. Desde el Sur, 13(2), e0018. https://doi.org/10.21142/des-1302-2021-0018

Wu, S.-M., Lei, Y.-Y., Miao, Q., Wang, J.-Q., Tian, Y.-F., & Xi, Z.-Y. (2023). The effect of modified observation-teaching-discussion pedagogy combined with microteaching on the critical thinking ability of nursing interns in China: A quasi-experimental study. *Frontiers in Education*, 8, 1276455. https://doi.org/10.3389/feduc.2023.1276455

Zhang, Q., Tang, H., & Xu, X. (2022). Analyzing collegiate critical thinking course effectiveness: Evidence from a quasi-experimental study in China. Thinking Skills and Creativity, 45, 101105. https://doi.org/10.1016/j.tsc.2022.101105