REVISTA INVECOM
ISSN en línea: 2739-0063

Preguntas socráticas y pensamiento crítico en estudiantes de secundaria

Socratic questioning and critical thinking in secondary school student

Recibido: 28/08/2026 - **Aceptado:** 11/11/2026

Beatriz Nancy Aquise Lagos
https://orcid.org/0009-0002-7043-8844
beatriz.aquise@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos. Lima, Perú

Abelardo Rodolfo Campana Concha
https://orcid.org/0000-0002-1098-9508
acampanac@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos. Lima, Perú

Jessica Paola Palacios Garay
https://orcid.org/0000-0002-2315-1683
jpalaciosg@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos. Lima, Perú

Elsa Esther Espinoza Garro
https://orcid.org/0000-0002-5068-6923
elsa.espinozag@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos. Lima, Perú

Cómo citar este artículo:
Aquise Lagos, B. N., Campana Concha, A. R., Palacios Garay, J. P., & Espinoza Garro, E. E. (2026). Preguntas socráticas y pensamiento crítico en estudiantes de secundaria. *Revista InveCom*, *6*(Núm. Especial 4), 1-9. https://doi.org/10.5281/zenodo.24503316

Resumen

La enseñanza secundaria continúa apoyándose, en gran medida, en preguntas cerradas, a pesar de la creciente necesidad de formar estudiantes capaces de analizar, argumentar y razonar con autonomía. Frente a esta brecha, el presente estudio evaluó el efecto de un programa de preguntas socráticas en el pensamiento crítico de estudiantes de tercero de secundaria de una institución pública de Comas, Lima. Para ello, se empleó un diseño cuasiexperimental con grupos intactos y mediciones de pretest y postest, en el que participaron 428 estudiantes distribuidos en dos grupos de 214. La intervención se desarrolló a lo largo de 16 sesiones de 90 minutos, mientras que el pensamiento crítico se midió mediante una rúbrica de 20 ítems organizada en cuatro dimensiones, con una alta consistencia interna ($\alpha = 0,947$). Los datos se analizaron mediante pruebas t, análisis de covarianza y cálculo de los tamaños del efecto. Los resultados fueron contundentes: el grupo experimental obtuvo una puntuación media de 53,06, frente a 30,00 del grupo control, lo que representa una diferencia de 23,06 puntos. Además, se observaron mejoras en las cuatro dimensiones evaluadas, siendo la dialógica la que presentó el mayor incremento. De manera consistente, el análisis de covarianza confirmó el efecto de la intervención tras ajustar por las diferencias iniciales entre los grupos ($F = 168,4$; $\eta^2p = 0,284$). En conjunto, los hallazgos respaldan la eficacia de las preguntas socráticas para promover el pensamiento crítico; no obstante, la magnitud del efecto observado hace necesario replicar el estudio e incorporar medidas de seguimiento que permitan valorar la permanencia de los resultados.

Palabras clave: preguntas socráticas, pensamiento crítico, diálogo pedagógico.

Abstract

Secondary education continues to rely heavily on closed-ended questions, despite the growing need to develop students capable of analyzing, arguing, and reasoning autonomously. To address this gap, this study evaluated the effect of a Socratic questioning program on the critical thinking skills of ninth-grade students at a public school in Comas, Lima. A quasi-experimental design with intact groups and pretest-posttest measurements was employed, involving 428 students divided into two groups of 214. The intervention consisted of 16 sessions, each lasting 90 minutes, while critical thinking was assessed using a 20-item rubric organized into four dimensions, which demonstrated high internal consistency ($\alpha = 0.947$). Data were analyzed using t-tests, analysis of covariance (ANCOVA), and effect size calculations. The results were striking: the experimental group achieved a mean score of 53.06, compared to 30.00 for the control group—a difference of 23.06 points. Furthermore, improvements were observed across all four dimensions, with the "dialogical" dimension showing the greatest increase. The analysis of covariance confirmed the intervention's effect after adjusting for initial differences between the groups ($F = 168.4$; $\eta^2p = 0.284$). Overall, these findings support the efficacy of Socratic questioning in promoting critical thinking; however, the magnitude of the observed effect warrants further replication of the study, as well as the inclusion of follow-up assessments to determine the long-term sustainability of these results.

Keywords: Socratic questioning, critical thinking, pedagogical dialogue.

Introducción

La escuela secundaria contemporánea se desarrolla en un entorno marcado por la saturación informativa, la circulación acelerada de datos y la polarización de las opiniones. En ese contexto, una enseñanza centrada en la memorización de contenidos resulta insuficiente. Formar estudiantes capaces de valorar la credibilidad de una fuente, identificar sesgos, construir argumentos sustentados y reconocer los límites de la propia posición dejó de ser un ideal curricular y se convirtió en una exigencia democrática.

Existe un amplio consenso pedagógico en torno a que el pensamiento crítico no se desarrolla de manera espontánea mediante la simple exposición a contenidos. Por el contrario, requiere intervenciones deliberadas que integren el modelado, la práctica guiada y la retroalimentación sostenida (Abrami et al., 2015; Bezanilla et al., 2019). Aun así, persiste una brecha entre ese consenso y las prácticas de aula, especialmente en instituciones públicas que enfrentan una alta carga curricular y una formación docente limitada en estrategias dialógicas.

En el aula ordinaria predomina un patrón de interacción conocido como iniciación-respuesta-evaluación. El docente formula una pregunta cuya respuesta ya conoce, un estudiante responde y el profesor confirma o corrige. Aunque este formato resulta útil para verificar información, restringe la elaboración conceptual, pues no exige justificación, contraste ni reformulación. Chin (2007) mostró que las preguntas que estimulan el pensamiento productivo se caracterizan por su apertura, su carácter de seguimiento y la disposición del docente a sostener el silencio mientras el estudiante piensa.

Por su parte, Mercer y Howe (2012) documentaron que el habla exploratoria, entendida como aquella en la que los participantes razonan de manera conjunta sobre ideas aún no resueltas, favorece aprendizajes conceptuales más profundos que la mera acumulación de intervenciones. En consecuencia, mejorar la calidad de las preguntas docentes constituye una de las palancas pedagógicas más accesibles y potentes de las que dispone una escuela.

Las preguntas socráticas representan la tradición más elaborada de este tipo de indagación. Su origen en la mayéutica platónica establece un principio epistémico según el cual el conocimiento no se transfiere, sino que se produce cuando el sujeto examina sus propias creencias y descubre sus inconsistencias. Overholser (1993) sistematizó el método en tres momentos, a saber, la indagación inductiva, la delimitación conceptual y el contraste analítico, y advirtió que su valor no reside en la secuencia de preguntas, sino en la calidad de la escucha que las acompaña.

Carey y Mullan (2004) precisaron una distinción decisiva entre el cuestionamiento socrático y el interrogatorio persuasivo. Mientras el primero busca el descubrimiento guiado por parte del propio estudiante, el segundo conduce hacia una conclusión predeterminada por el docente. Asimismo, Paul y Elder (2006) tradujeron ese legado en una tipología operativa de seis clases de preguntas, relacionadas con la clarificación, los supuestos, las razones y evidencias, los puntos de vista, las consecuencias y la metacognición, ampliamente adoptada en el diseño de secuencias didácticas.

El fundamento psicológico de esta estrategia se encuentra en la tradición sociocultural. La pregunta actúa como un instrumento de mediación que opera en la zona de desarrollo próximo, pues no resuelve el problema por

el estudiante, pero sostiene su actividad reflexiva hasta que logra un nivel de elaboración que no alcanzaría por sí solo (Mercer & Howe, 2012). Lipman (2003) extendió ese principio a la dinámica del aula mediante la noción de comunidad de indagación, en la que el grupo construye colectivamente criterios de validez y aprende a someter toda afirmación a examen.

De igual manera, Nystrand et al. (1997) aportaron evidencia empírica al demostrar que las preguntas auténticas, carentes de una respuesta predeterminada, incrementan la profundidad del razonamiento en comparación con la recitación. Kuhn (2019) complementó esta perspectiva al situar el pensamiento crítico en el terreno del discurso argumentativo, pues se piensa críticamente cuando se participa en intercambios en los que las razones deben ser articuladas, defendidas y revisadas.

El constructo de pensamiento crítico es multidimensional, lo cual condiciona su evaluación. El consenso Delphi lo definió como un juicio autorregulado y con propósito, que se manifiesta en la interpretación, el análisis, la evaluación, la inferencia, la explicación y la autorregulación (Facione, 1990). Ennis (1993), a su vez, distinguió las dimensiones lógica, criterial y pragmática, y subrayó que la disposición a preguntar y a buscar razones es tan importante como la destreza para evaluar argumentos.

La taxonomía revisada de procesos cognitivos permite situar esta progresión desde el recuerdo y la comprensión hacia el análisis, la evaluación y la creación (Anderson & Krathwohl, 2001). Estos son los niveles en los que se ubica el pensamiento crítico y que exigen tareas de elaboración, no de reproducción. Moore (2011) añadió una advertencia relevante para el diseño de intervenciones, ya que el pensamiento crítico se ejerce sobre contenidos disciplinares; por ello, enseñarlo como una técnica genérica produce transferencias débiles. Villarini (2003), desde el contexto latinoamericano, vinculó la competencia crítica con la ética y la convivencia democrática, lo que refuerza su carácter formativo y no solo instrumental.

La evidencia empírica sobre intervenciones es abundante, aunque desigual. El metaanálisis de Abrami et al. (2015) estimó un efecto medio de 0,30 desviaciones estándar para los programas de enseñanza del pensamiento crítico, con una variabilidad considerable según el tipo de intervención y la calidad del instrumento. Tiruneh et al. (2014) encontraron que los programas integrados en el currículo superan a los cursos aislados, mientras que Bezanilla et al. (2019) documentaron que los docentes identifican la discusión y el aprendizaje basado en problemas como las metodologías más eficaces, aunque reconocen una formación insuficiente para conducirlas.

Los estudios específicamente socráticos son menos numerosos. Yang et al. (2005) verificaron mejoras en foros asincrónicos, y Bozer y Korkmaz (2025) reportaron efectos positivos en habilidades de orden superior en estudiantes de secundaria. Sin embargo, persiste una brecha importante, pues escasean los estudios cuasiexperimentales con grupos de control comparables, tamaños muestrales suficientes y reporte de tamaños del efecto con intervalos de confianza en contextos escolares públicos de América Latina.

Por ello, el objetivo de este estudio fue determinar el efecto de un programa de preguntas socráticas sobre el pensamiento crítico en estudiantes de tercer grado de secundaria de una institución educativa pública del distrito de Comas, Lima, durante el año 2025. Los objetivos específicos consistieron en estimar el efecto sobre las dimensiones sustantiva, lógica, dialógica y contextual, así como en evaluar la robustez del efecto tras ajustar por la medición inicial.

La hipótesis general sostuvo que el programa incrementa significativamente el pensamiento crítico en comparación con el grupo de control. Las hipótesis específicas replicaron esa expectativa en cada una de las dimensiones evaluadas. El aporte de este estudio es triple, ya que ofrece una estimación con tamaño del efecto e intervalos de confianza en un contexto escasamente documentado, describe con precisión operativa una secuencia de dieciséis sesiones replicable por otros docentes y discute críticamente la magnitud del efecto frente a los promedios metaanalíticos.

Metodología

Se realizó una investigación con enfoque cuantitativo, de tipo aplicada y diseño cuasiexperimental de medición pretest y postest con grupo de control no equivalente (Hernández-Sampieri & Mendoza, 2018). El estudio respondió al esquema en el que el grupo experimental recibió la intervención entre ambas mediciones, mientras que el grupo de control continuó con la enseñanza habitual, según la secuencia GE O1–X–O2 y GC O1–O2. La asignación no fue aleatoria, debido a que los grupos ya estaban constituidos por secciones intactas, condición habitual en la investigación educativa escolar. Para compensar esta limitación, se verificó la equivalencia inicial de los grupos y se ajustó el efecto por la medición pretest mediante un análisis de covarianza.

La población estuvo conformada por 612 estudiantes matriculados en el tercer grado de secundaria de una institución educativa pública del distrito de Comas, en los turnos de mañana y tarde. Se empleó un muestreo

no probabilístico por conveniencia con grupos intactos, criterio frecuente cuando la intervención debe desarrollarse dentro del horario escolar sin alterar la organización institucional.

La muestra final quedó constituida por 428 estudiantes, de los cuales 214 pertenecieron al grupo experimental, distribuidos en cuatro secciones, y 214 al grupo de control, también organizados en cuatro secciones. Esta muestra representó el 69,9% de la población. Se incluyó a los estudiantes con matrícula vigente, asentimiento informado y consentimiento firmado por el apoderado, así como a quienes completaron ambas mediciones. Por el contrario, se excluyó a los estudiantes que registraron menos del 80% de asistencia a las sesiones o que dejaron ítems en blanco.

El programa de intervención se organizó en 16 sesiones de 90 minutos, aplicadas durante ocho semanas con una frecuencia de dos sesiones semanales. Cada sesión mantuvo una estructura constante, compuesta por la activación de saberes previos mediante una situación problemática, una secuencia de preguntas socráticas organizada según la tipología de Paul y Elder (2006), el trabajo en tríadas con los roles de hablante, cuestionador y observador, y un cierre metacognitivo en el que los estudiantes registraron qué pregunta los había obligado a cambiar de posición. Los docentes responsables recibieron 20 horas de formación previa en conducción de diálogo y manejo de silencios. Además, su desempeño se monitoreó mediante una lista de cotejo de fidelidad aplicada en 14 de las 16 sesiones, con un cumplimiento medio del 92%.

El pensamiento crítico se evaluó mediante una rúbrica de 20 ítems distribuidos en cuatro dimensiones de cinco ítems cada una. La dimensión sustantiva abordó la revisión de las propias opiniones y la jerarquización de fuentes; la lógica, la validez de las inferencias y la detección de falacias; la dialógica, la escucha activa, la apertura a la refutación y la valoración de perspectivas; y la contextual, la transferencia a problemas del entorno.

Cada ítem se puntuó en cuatro niveles, denominados inicio, proceso, logro previsto y logro destacado, con valores de 1 a 4. En consecuencia, el puntaje total osciló entre 20 y 80 puntos, mientras que cada dimensión alcanzó un rango de 5 a 20 puntos. La validez de contenido se estableció mediante el juicio de cinco expertos con grado de doctor, obteniéndose un coeficiente V de Aiken de 0,93 para el conjunto. La confiabilidad se estimó en una prueba piloto con 30 estudiantes de características similares, en la que se obtuvo un alfa de Cronbach de 0,947 para la escala total y valores entre 0,86 y 0,91 en las dimensiones, lo que evidencia una consistencia interna alta (George & Mallery, 2003).

Los datos se procesaron con IBM SPSS Statistics versión 28. Previamente, se verificó la normalidad mediante la prueba de Shapiro-Wilk y la homogeneidad de varianzas con la prueba de Levene. Las comparaciones entre grupos se realizaron con la prueba t de Student para muestras independientes, mientras que el efecto se ajustó mediante un análisis de covarianza (ANCOVA), empleando el pretest como covariable.

Los tamaños del efecto se estimaron con la d de Cohen y sus intervalos de confianza del 95%, interpretados de acuerdo con los criterios convencionales (Cohen, 1988). Asimismo, las diferencias en la distribución de los niveles de logro se examinaron mediante la prueba chi-cuadrado y la V de Cramer. Se adoptó un nivel de significancia de 0,05 y se verificó la robustez de los resultados al excluir los casos extremos.

Resultados

Los resultados se presentan en cuatro apartados. Primero, se examina la equivalencia inicial de los grupos; luego, se describen los cambios en los niveles de logro, las diferencias postest en las dimensiones del pensamiento crítico y la robustez del efecto tras el ajuste estadístico y la exclusión de casos extremos.

Equivalencia inicial de los grupos

Los grupos resultaron equivalentes en la medición inicial. En el pretest, el grupo experimental obtuvo una media de 24,86 puntos (DE = 6,12), mientras que el grupo de control alcanzó 25,31 puntos (DE = 6,40), sin diferencias estadísticamente significativas entre ambos (t = −0,73t; gl = 426; p = 0,466; d = 0,07; IC 95% [−0,26; 0,12]). La prueba de Levene confirmó la homogeneidad de varianzas (F = 1,87; p = 0,172).

Asimismo, la distribución de los niveles de logro fue semejante en ambos grupos. Aproximadamente siete de cada diez estudiantes se ubicaron en el nivel de inicio antes de la intervención, lo que respalda la comparabilidad inicial de las condiciones experimentales (Tabla 1). Esta equivalencia constituye el supuesto que habilita la interpretación de los cambios posteriores.

Tabla 1
Niveles de logro en pensamiento crítico antes y después de la intervención

Nivel de logro	Pretest GC n (%)	Pretest GE n (%)	Postest GC n (%)	Postest GE n (%)
Inicio	153 (71,5)	163 (76,2)	138 (64,5)	9 (4,2)
Proceso	53 (24,8)	43 (20,1)	61 (28,5)	26 (12,1)
Logro previsto	8 (3,7)	8 (3,7)	13 (6,1)	51 (23,8)
Logro destacado	0 (0,0)	0 (0,0)	2 (0,9)	128 (59,8)
Total	**214 (100)**	**214 (100)**	**214 (100)**	**214 (100)**

Nota. GC = grupo de control; GE = grupo experimental | **Fuente.** Elaboración propia

Cambios en los niveles de logro

Tras la intervención, la distribución del grupo experimental se modificó de manera sustancial. El 59,8% de sus integrantes alcanzó el nivel de logro destacado y solo el 4,2% permaneció en el nivel de inicio. En contraste, el grupo de control continuó concentrado en el nivel de inicio, con el 64,5% de los estudiantes, y únicamente el 0,9% alcanzó el nivel de logro destacado.

La diferencia entre los grupos en la distribución postest fue estadísticamente significativa ($\chi^2(3) = 272,0$; $p < 0,001$; V de Cramer = 0,80), lo que evidencia una asociación fuerte entre la condición experimental y el nivel de logro alcanzado. El grupo de control presentó una mejora marginal, atribuible a la enseñanza habitual; no obstante, esta resultó insuficiente para modificar su posición relativa frente al grupo experimental.

Diferencias en pensamiento crítico

Las comparaciones de medias confirmaron la hipótesis general y las cuatro hipótesis específicas (Tabla 2). En el puntaje global, el grupo experimental alcanzó una media de 53,06 puntos, frente a 30,00 puntos del grupo de control, con una diferencia de 23,06 puntos (IC 95 % [20,28; 25,84]; $t = 16,27$; $gl = 426$; $p < 0,001$; $d = 1,57$).

Las cuatro dimensiones presentaron diferencias significativas y tamaños del efecto grandes. La dimensión dialógica registró el mayor incremento, con una diferencia de 6,57 puntos ($d = 1,51$), seguida de la contextual, con 6,32 puntos ($d = 1,50$); la sustantiva, con 5,41 puntos ($d = 1,27$); y la lógica, con 4,76 puntos ($d = 1,22$). Este patrón resulta coherente con la naturaleza conversacional de la intervención, orientada a promover la escucha, el contraste de perspectivas y la revisión de las propias posiciones.

Tabla 2
Comparación de medias postest por dimensión

Dimensión	GE M (DE)	GC M (DE)	Diferencia [IC 95 %]	t	p	d
Sustantiva	14,62 (4,10)	9,21 (4,40)	5,41 [4,60; 6,22]	13,16	< 0,001	1,27
Lógica	12,88 (3,80)	8,12 (4,00)	4,76 [4,02; 5,50]	12,63	< 0,001	1,22
Dialógica	13,42 (4,50)	6,85 (4,20)	6,57 [5,74; 7,40]	15,61	< 0,001	1,51
Contextual	12,14 (4,30)	5,82 (4,10)	6,32 [5,53; 7,11]	15,57	< 0,001	1,50
Pensamiento crítico global	53,06 (14,20)	30,00 (15,10)	23,06 [20,28; 25,84]	16,27	< 0,001	1,57

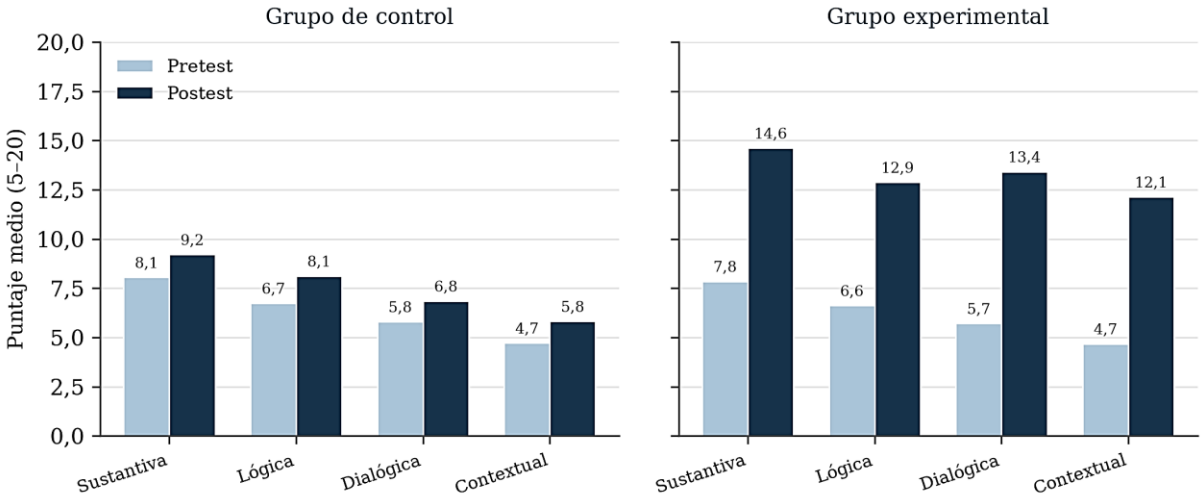
Nota. GE = grupo experimental; GC = grupo de control; IC 95 % = intervalo de confianza de la diferencia de medias; d = d de Cohen | **Fuente.** Elaboración propia

La Figura 1 permite observar el patrón de cambio por dimensión en cada grupo. Mientras que el grupo de control experimentó incrementos pequeños y relativamente homogéneos, el grupo experimental mostró una ganancia pronunciada en todas las dimensiones. No obstante, la dimensión lógica, considerada la más exigente desde el punto de vista cognitivo, presentó el menor avance relativo.

Este hallazgo sugiere que el diálogo socrático fortalece con rapidez la disposición de los estudiantes para examinar sus propios supuestos y considerar perspectivas ajenas. Sin embargo, la evaluación formal de

inferencias parece requerir una práctica adicional, especialmente mediante ejercicios de lógica explícita que permitan consolidar dicha habilidad.

Figura 1
Medias pretest y postest por dimensión y grupo (N = 428)



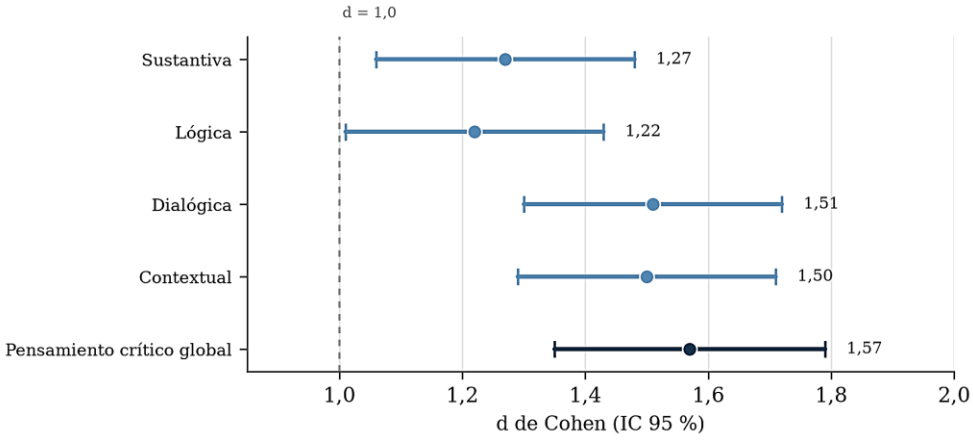
Nota. Las barras claras corresponden al pretest y las barras oscuras al postest. Cada dimensión se puntúa entre 5 y 20 | **Fuente.** Elaboración propia

Robustez del efecto

El análisis de covarianza, que controló el puntaje inicial, confirmó el efecto de la intervención sobre el pensamiento crítico global ($F(1, 425) = 168{,}4$; $p < 0{,}001$; $\eta^2p = 0{,}284$). Las medias ajustadas fueron de 52,41 puntos para el grupo experimental y de 30,65 puntos para el grupo de control. Como era esperable, el pretest también explicó una porción significativa de la varianza ($F(1, 425) = 41{,}7$; $p < 0{,}001$; $\eta^2p = 0{,}089$), lo que confirma la pertinencia del ajuste estadístico.

La Figura 2 resume la magnitud de los efectos con sus intervalos de confianza. Todos los intervalos se ubicaron por encima de 1,0, umbral que corresponde a un efecto muy grande según los criterios convencionales (Cohen, 1988).

Figura 2
Tamaños del efecto (d de Cohen) con intervalos de confianza del 95 %



Nota. La línea vertical punteada indica d = 1,0 | **Fuente.** Elaboración propia

Finalmente, se examinó la robustez de los hallazgos. La exclusión de los seis casos con puntajes extremos en el postest, tres por grupo y definidos como valores fuera del rango intercuartílico por 1,5 veces, mantuvo la diferencia global en 22,84 puntos ($t = 15,92$; $p < 0,001$; $d = 1,54$). Asimismo, el análisis por dimensiones conservó el mismo orden de magnitud.

Por otra parte, el cumplimiento de la fidelidad de la intervención, registrado en 14 de las 16 sesiones, no mostró una asociación significativa con el puntaje alcanzado ($r = 0,08$; $p = 0,098$). Este resultado sugiere que el efecto observado no depende de variaciones menores en la aplicación del programa.

Discusión

Los resultados confirman que un programa estructurado de preguntas socráticas incrementa el pensamiento crítico de estudiantes de secundaria, con un efecto que se mantiene al ajustar por la medición inicial y al excluir los casos extremos. La dirección del hallazgo converge con la evidencia disponible, pues Bozer y Korkmaz (2025) reportaron mejoras en habilidades de orden superior tras una intervención socrática en secundaria, mientras que Yang et al. (2005) documentaron ganancias equivalentes en entornos de discusión asincrónica. Asimismo, el resultado coincide con la conclusión de Abrami et al. (2015), según la cual las intervenciones prolongadas con práctica guiada superan a las exposiciones breves. La novedad del presente estudio no reside en la dirección del efecto, ya anticipada por la literatura, sino en su estimación con intervalos de confianza en un contexto escolar público peruano escasamente documentado.

La magnitud del efecto, en cambio, exige una lectura cautelosa. Con $d = 1,57$, el efecto observado multiplica por cinco el promedio metaanalítico de 0,30 reportado por Abrami et al. (2015) y excede con holgura los valores habituales de las intervenciones escolares. Por ello, son plausibles tres explicaciones que no se excluyen entre sí. La primera se relaciona con la alineación entre el instrumento y la intervención, ya que la rúbrica evalúa precisamente las conductas que el programa entrena, tales como clarificar, justificar, contrastar perspectivas y transferir. En consecuencia, parte del efecto puede reflejar la especificidad de la medida, más que una transferencia general del pensamiento crítico.

La segunda explicación corresponde a la intensidad de la intervención. Dieciséis sesiones de noventa minutos equivalen a 24 horas de práctica deliberada, una carga temporal muy superior a la mediana de las intervenciones incluidas en los metaanálisis. La tercera explicación se vincula con el punto de partida de los estudiantes. Dado que dos tercios de la muestra se ubicaron inicialmente en el nivel de inicio, el margen de mejora era amplio y los enunciados de la rúbrica podían resultar poco familiares, lo que pudo inflar la ganancia observada en el postest. Tiruneh et al. (2014) advirtieron que los efectos grandes en este campo suelen disminuir cuando se emplean pruebas estandarizadas externas, advertencia que debe ponderarse al interpretar los resultados.

El patrón por dimensiones aporta un argumento teórico relevante. La mayor ganancia se produjo en la dimensión dialógica y la menor en la lógica, un orden que no parece explicarse por el azar, sino por el mecanismo de la intervención. Desde la perspectiva sociocultural, la pregunta opera en la zona de desarrollo próximo y su efecto inmediato se manifiesta en la participación. El estudiante aprende a escuchar, esperar su turno argumentativo y reformular, conductas que el formato de trabajo en tríadas entrena de manera directa (Mercer & Howe, 2012).

La evaluación formal de inferencias, en cambio, requiere un componente de lógica explícita que el diálogo solo proporciona de forma indirecta, de allí su avance comparativamente menor. Este resultado dialoga con la tesis de Kuhn (2019), para quien el pensamiento crítico es ante todo una práctica discursiva, y matiza la lectura puramente cognitiva de Ennis (1993). La destreza lógica no se adquiere únicamente mediante la inmersión conversacional, sino que necesita instrucción específica y oportunidades de práctica estructurada.

La comparación con la teoría de Lipman (2003) refuerza esta interpretación. La comunidad de indagación no se limita a conversar, sino que instala procedimientos de validación colectiva que transforman la clase en un espacio donde las razones se examinan públicamente. Los estándares intelectuales formulados por Paul y Elder (2020), relacionados con la claridad, la exactitud, la precisión, la pertinencia, la profundidad y la amplitud, operaron como un vocabulario compartido en el aula. Así, los estudiantes aprendieron a nombrar la deficiencia de una respuesta antes de corregirla. Los cierres metacognitivos del programa, en los que cada estudiante identificaba la pregunta que lo había obligado a cambiar de posición, funcionaron como dispositivos de objetivación de ese proceso y explican, en parte, por qué la dimensión sustantiva, asociada a la revisión de las propias opiniones, también registró ganancias apreciables ($d = 1,27$).

Por su parte, Moore (2011) permitiría anticipar una objeción relevante. Al entrenar el cuestionamiento con contenidos genéricos y no con las epistemologías propias de cada disciplina, la transferencia a la argumentación histórica o científica podría resultar limitada. El diseño del programa intentó mitigar esta limitación al trabajar con

situaciones problemáticas del entorno; sin embargo, el estudio no evaluó la transferencia disciplinar, la cual queda como una tarea pendiente para investigaciones futuras.

En el plano práctico, los hallazgos sustentan tres decisiones. Primero, la formación docente en formulación de preguntas y conducción de silencios debería ocupar un lugar central en los programas de desarrollo profesional, en línea con lo que los propios docentes identifican como su principal carencia (Bezanilla et al., 2019). Segundo, la evaluación del pensamiento crítico requiere rúbricas con dimensiones explícitas y niveles observables, dado que las pruebas de opción múltiple no capturan la calidad de la argumentación. Tercero, la sostenibilidad exige institucionalizar la estrategia en la programación curricular, en lugar de confiar en intervenciones aisladas, tal como sugiere la evidencia sobre programas integrados (Tiruneh et al., 2014).

Por último, es importante señalar que el estudio presenta limitaciones que acotan el alcance de sus conclusiones. La asignación no aleatoria de grupos intactos introduce un riesgo de sesgo debido a diferencias no observadas entre los participantes. Aunque la equivalencia inicial y el ajuste mediante covarianza reducen ese riesgo, no lo eliminan por completo. Asimismo, la intervención fue conducida por docentes que conocían la hipótesis del estudio, lo que pudo generar expectativas favorables e influir en su comportamiento dentro del aula.

El instrumento, de elaboración propia, cuenta con validez de contenido y una alta confiabilidad; sin embargo, no se ha demostrado su equivalencia respecto de pruebas estandarizadas internacionales. Tampoco se realizó una medición de seguimiento, por lo que se desconoce si las ganancias obtenidas se mantienen en el tiempo. La literatura sugiere que los efectos de este tipo de programas tienden a disminuir cuando no se garantiza una práctica sostenida.

La muestra, aunque amplia para el contexto de estudio, proviene de una sola institución educativa y de un distrito específico, lo que limita la generalización de los resultados. No obstante, entre las fortalezas del estudio destacan el tamaño muestral, la existencia de un grupo de control comparable, el reporte de tamaños del efecto con intervalos de confianza, el análisis de covarianza y el registro de la fidelidad de la implementación.

Conclusiones

Un programa de dieciséis sesiones basado en preguntas socráticas incrementó significativamente el pensamiento crítico de estudiantes de tercero de secundaria de una institución pública de Comas, con una diferencia de 23,06 puntos sobre el grupo de control ($t = 16,27$; $gl = 426$; $p < 0,001$; $d = 1,57$) que se mantuvo al ajustar por la medición inicial ($F = 168,4$; $\eta^2p = 0,284$) y al excluir casos extremos.

El efecto se verificó en las cuatro dimensiones evaluadas, con el mayor incremento en la dialógica y el menor en la lógica, y se acompañó de una redistribución sustantiva de los niveles de logro: casi seis de cada diez estudiantes del grupo experimental alcanzaron el nivel destacado, frente a menos de uno por ciento en el grupo de control.

Estos resultados respaldan la incorporación de la indagación socrática como estrategia regular en la secundaria pública, siempre que se acompañe de formación docente, rúbricas explícitas y continuidad en la programación. La magnitud del efecto, superior a los promedios meta-analíticos, invita a replicar el estudio con pruebas externas, medición de seguimiento y diseño aleatorizado por conglomerados, así como a examinar la transferencia a dominios disciplinares específicos y el papel de la metacognición como mediadora. La pregunta bien formulada sigue siendo, en contextos de información saturada, uno de los instrumentos más económicos y potentes para formar ciudadanos capaces de pensar por sí mismos.

Referencias

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

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Bezanilla, M. J., Fernández-Nogueira, D., Poblete, M., & Galindo-Domínguez, H. (2019). Methodologies for teaching-learning critical thinking in higher education: The teacher's view. *Thinking Skills and Creativity, 33*, 100584. https://doi.org/10.1016/j.tsc.2019.100584

Bozer Özsaraç, F., & Korkmaz, Ö. (2025). The effect of Socratic questioning on secondary school students' higher-order thinking. *International Journal of Modern Education Studies, 9*(2), 1–22. https://doi.org/10.51383/ijonmes.2025.424

Carey, T. A., & Mullan, R. J. (2004). What is Socratic questioning? *Psychotherapy: Theory, Research, Practice, Training, 41*(3), 217–226. https://doi.org/10.1037/0033-3204.41.3.217

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching, 44*(6), 815–843. https://doi.org/10.1002/tea.20171
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2da ed.). Lawrence Erlbaum.
Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice, 32*(3), 179–186. https://doi.org/10.1080/00405849309543594
Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction* [The Delphi Report]. American Philosophical Association.
George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4ta ed.). Allyn & Bacon.
Hernández-Sampieri, R., & Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
Kuhn, D. (2019). Critical thinking as discourse. *Human Development, 62*(3), 146–164. https://doi.org/10.1159/000500171
Lipman, M. (2003). *Thinking in education* (2da ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840272
Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction, 1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001
Moore, T. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development, 30*(3), 261–274. https://doi.org/10.1080/07294360.2010.501328
Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. Teachers College Press.
Overholser, J. C. (1993). Elements of the Socratic method: I. Systematic questioning. *Psychotherapy: Theory, Research, Practice, Training, 30*(1), 67–74. https://doi.org/10.1037/0033-3204.30.1.67
Paul, R., & Elder, L. (2006). *The art of Socratic questioning*. Foundation for Critical Thinking.
Paul, R., & Elder, L. (2020). *Critical thinking: Tools for taking charge of your learning and your life* (4ta ed.). Foundation for Critical Thinking.
Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies, 4*(1), 1–17. https://doi.org/10.5539/hes.v4n1p1
Villarini, Á. R. (2003). Teoría y pedagogía del pensamiento crítico. *Perspectivas Psicológicas, 3–4*, 35–42. https://pepsic.bvsalud.org/pdf/pp/v3-4/v3-4a04.pdf
Yang, Y.-T. C., Newby, T. J., & Bill, R. L. (2005). Using Socratic questioning to promote critical thinking skills through asynchronous discussion forums in distance learning environments. *The American Journal of Distance Education, 19*(3), 163–181. https://doi.org/10.1207/s15389286ajde1903_4

CONTRIBUCIÓN DE LOS AUTORES:

Conceptualización: Beatriz Nancy Aquise Lagos.
Metodología: Beatriz Nancy Aquise Lagos, Abelardo Rodolfo Campana Concha y Elsa Esther Espinoza Garro.
Investigación: Beatriz Nancy Aquise Lagos y Jessica Paola Palacios Garay.
Análisis formal: Abelardo Rodolfo Campana Concha y Elsa Esther Espinoza Garro.
Validación: Abelardo Rodolfo Campana Concha.
Curación de datos: Jessica Paola Palacios Garay.
Visualización: Jessica Paola Palacios Garay.
Administración del proyecto: Beatriz Nancy Aquise Lagos.
Supervisión: Abelardo Rodolfo Campana Concha.
Redacción – borrador original: Beatriz Nancy Aquise Lagos.
Redacción – revisión y edición: Abelardo Rodolfo Campana Concha, Jessica Paola Palacios Garay y Elsa Esther Espinoza Garro.