

Efectos de una estrategia de desarrollo profesional en las creencias y habilidades de argumentación científica en docentes universitarios

Effects of a Professional Development Strategy on University Teachers' Beliefs and Scientific Argumentation Skills

Recibido: 15/9/2025 / Revisado: 22/9/2025 / Aprobado: 23/11/2025

Alejandra Martínez-Morales^a 

Milany Gómez Betancur^b 

^{a,b} Universidad Antonio Nariño, Bogotá, Colombia.

Autora de correspondencia: almartinez54@uan.edu.co

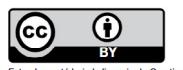
Resumen

Este artículo analiza el impacto de la estrategia de desarrollo profesional docente mediada por tecnología (FAEUC) en el fortalecimiento de las creencias, conocimientos y prácticas argumentativas de cinco docentes universitarios de ciencias, mediante un diseño pretest–posttest sin grupo control. La estrategia integró modelación de argumentos, análisis de casos y elaboración de tareas argumentativas apoyadas con herramientas digitales. Los resultados del Cuestionario de Evaluación del Fomento y Uso de la Argumentación (CUEFOAR) y el *Argument Test* (AT) evidenciaron incrementos significativos en la valoración de la enseñanza de la estructura argumentativa, el fomento de habilidades analíticas y la capacidad para distinguir argumentos bien sustentados y formular cuestionamientos más precisos. La lectura conjunta de los instrumentos mostró coherencia entre lo que los docentes declaran y lo que efectivamente hacen al analizar argumentos, con matices según disciplina y experiencia. A partir de estos hallazgos, se recomienda explicitar la estructura de Toulmin, fortalecer el trabajo con refutaciones y emplear rúbricas analíticas por su utilidad para evaluar de manera consistente los componentes de la argumentación. Aunque la muestra es reducida, los resultados indican un efecto positivo de la estrategia y orientan la necesidad de consolidar criterios comunes y un banco de tareas por disciplina.

Palabras clave: argumentación científica, desarrollo profesional, evaluación, intervención educativa

ISSN (en línea): 1814-4152 / Sitio web: <http://cuaderno.pucmm.edu.do>

CÓMO CITAR: Martínez-Morales, A. y Gómez, M. (2026). Efectos de una estrategia de desarrollo profesional en las creencias y habilidades de argumentación científica en docentes universitarios. *Cuaderno de Pedagogía Universitaria*, e702



Esta obra está bajo la licencia de Creative Commons Atribución 4.0 Internacional

Abstract

This article examines the impact of a technology-supported professional development strategy (FAEUC) on strengthening the beliefs, knowledge, and practices related to scientific argumentation of five university science instructors, using a pretest–posttest design without a control group. The strategy incorporated argument modeling, case analysis, and the design of argumentative tasks supported by digital tools. Findings from the Questionnaire for Evaluating the Promotion and Use of Argumentation (CUEFOAR) and the Argument Test (AT) showed marked improvements in teachers' valuation of explicit argument-structure instruction, the promotion of analytical skills, and their ability to distinguish well-supported arguments and formulate more precise challenges. A joint reading of both instruments revealed coherence between what instructors report doing and how they actually analyze arguments, with nuances related to disciplinary background and teaching experience. Based on these results, it is recommended to make Toulmin's structure explicit in instruction, strengthen work with refutations, and use analytic rubrics given their value for consistently assessing components of argumentation. Although the sample is small, the findings indicate a positive effect of the strategy and point to the need to consolidate common criteria and develop a discipline-specific bank of tasks.

Keywords: scientific argumentation, professional development, evaluation, educational intervention

Introducción

La argumentación científica es una práctica central de las ciencias y un medio para desarrollar habilidades de pensamiento crítico, alfabetización científica y comunicación rigurosa en educación superior (Giri y Paily, 2020; Jiménez-Aleixandre y Erduran, 2007; Li et al., 2022; Osborne et al., 2004). En la educación universitaria colombiana, persiste el reto de que el profesorado integre estas prácticas de forma sistemática, lo cual limita el desarrollo de competencias críticas en el estudiantado (Benarroch et al., 2024). Ejemplos como la formulación de explicaciones basadas en evidencia, la evaluación crítica de afirmaciones en laboratorio y la justificación de decisiones metodológicas aún resultan poco frecuentes en cursos universitarios (Archila et al., 2022; Faize et al., 2018).

La investigación en didáctica de las ciencias ha evidenciado que, aunque las creencias sobre la importancia de argumentar suelen ser positivas, persisten limitaciones en los conocimientos didácticos necesarios para enseñarla y en la puesta en práctica de actividades que permitan evaluarla en el aula (Erduran y Jiménez-Aleixandre, 2007; Osborne et al., 2013; Wess et al., 2023). De forma coincidente, los estudios sobre desarrollo profesional destacan la necesidad de integrar dimensiones conceptuales, epistémicas y didácticas mediante estrategias sostenidas que transformen de manera integrada las creencias, conocimientos y prácticas del profesorado (Darling-Hammond et al., 2017; Zohar y Resnick, 2021).

En este contexto, este estudio indaga cómo una estrategia de desarrollo profesional docente puede fortalecer las creencias, conocimientos y prácticas argumentativas del profesorado universitario de ciencias, al tiempo que se atiende una necesidad identificada en la literatura: la falta de formación sistemática para integrar la argumentación como práctica científica en la educación superior (Benarroch et al., 2024; Martínez-Morales y Briceño-Martínez, 2025; Osborne et al., 2013; Wess et al., 2023). Esta necesidad formativa sustenta el propósito de esta investigación, orientada a examinar cómo una estrategia de desarrollo profesional puede apoyar al profesorado en la integración sistemática de la argumentación científica.

La estrategia Fomento de la Argumentación en la Educación Universitaria de Ciencias (FAEUC) busca responder a esta brecha mediante un enfoque formativo que promueve el razonamiento crítico y la enseñanza basada en evidencias, además de que se implementa en modalidad virtual y asincrónica con docentes universitarios de física, química y biología en Colombia. Su intención pedagógica es fortalecer las competencias para promover y evaluar la argumentación en clase, con base en fundamentos conceptuales y herramientas didácticas derivadas de Plantin (1998, 2009), Toulmin (2007), Walton (1996) y en el Conocimiento Pedagógico del Contenido de la Argumentación (PCK-A) (Osborne et al., 2013). De ahí surge el objetivo de explorar su impacto en la argumentación científica del profesorado universitario.

Los apartados posteriores presentan el marco teórico, la descripción de la estrategia FAEUC, la metodología del estudio y los resultados pre–post, seguidos de una discusión sobre sus aportes y proyecciones.

Fundamentación teórica

En la educación en ciencias, la argumentación no es un complemento retórico, sino una práctica central mediante la cual se construyen, justifican y reconocen conocimientos (Felton et al., 2022; Li et al., 2022; Jiménez-Aleixandre y Erduran, 2007). Desde un enfoque sociodiscursivo, Plantin (2009) concibe la argumentación como una actividad contextualizada que articula dimensiones lógica, retórica y pragmática, por lo que, al trasladarse al aula universitaria, exige situaciones dialógicas donde el estudiantado justifique sus afirmaciones con evidencias, considere objeciones y revise sus posturas. Estas situaciones incluyen prácticas como la defensa oral de procedimientos experimentales, la revisión crítica de conclusiones en informes de laboratorio y la comparación de modelos explicativos entre pares (Archila et al., 2022; Faize et al., 2018). Estas prácticas demandan del profesorado conocimientos que permitan diseñar tareas, normas y herramientas adecuadas (Osborne et al., 2004; Sandoval y Millwood, 2007; Wess et al., 2023).

El Patrón de Argumentación de Toulmin (TAP, por sus siglas en inglés) brinda un marco analítico y didáctico que delimita componentes —datos, garantías, respaldos, calificadores y refutaciones— (Toulmin, 2007). En didáctica de las ciencias, su

uso eleva la calidad epistémica del discurso, hace visibles los vínculos evidencia–conclusión y fortalece habilidades para rebatir argumentos (Osborne et al., 2013; Wess et al., 2023). Así, en términos de progresión, el estudiantado desarrolla niveles epistémicos que reflejan su dominio en el empleo de evidencias y justificaciones; por ejemplo, progresan desde respuestas basadas en descripciones hacia argumentos que articulan datos con garantías explícitas y refutaciones pertinentes. A su vez, el profesorado que domina estos componentes dispone de criterios de evaluación alineados con la práctica científica (Kelly y Takao, 2002; Zohar, 2007; Zohar y Resnick, 2021). En este estudio, el TAP se utiliza principalmente como lente analítico para caracterizar este tipo de progresión.

Para que la estructura argumentativa trascienda lo declarativo y se exprese en interacciones auténticas de aula, es útil articular tres enfoques complementarios. En primer lugar, la organización de clases como comunidades de indagación permite establecer criterios de explicitación de datos, diálogo respetuoso y contraste crítico de ideas (Osborne et al., 2013; Wess et al., 2023). Por su parte, el modelo dialógico de Walton (1996) aporta esquemas de interacción que orientan la conducción de discusiones y ofrecen criterios para evaluar su progreso. Finalmente, enfoques como la Argument-Based Inquiry (ABI) y los estudios de debates contribuyen a la construcción colectiva de argumentos apoyados en datos, la consideración de perspectivas alternativas y la elaboración de refutaciones con criterios explícitos (Sampson y Blanchard, 2012; Jiménez-Aleixandre y Erduran, 2007). En conjunto, estos referentes operativizan el TAP en situaciones didácticas que promueven el uso de evidencias y las oportunidades de diálogo.

Por lo anterior, sostener y escalar estas prácticas requiere atender tres dimensiones clave: lo epistémico (uso explícito de evidencias y garantías), lo discursivo (calidad del diálogo y de las refutaciones) y lo didáctico (selección de tareas y apoyos para el análisis). Esto implica integrar el Conocimiento Pedagógico del Contenido de la Argumentación con el saber disciplinar y el conocimiento del estudiantado, seleccionar tareas abundantes en evidencia, diseñar apoyos para el diálogo, anticipar dificultades comunes y ofrecer retroalimentación formativa (Adúriz-Bravo, 2014; Osborne et al., 2013; Zohar y Nemet, 2002). En coherencia con la evidencia sobre desarrollo profesional, las iniciativas más eficaces se centran en el contenido, con práctica auténtica, materiales de apoyo, retroalimentación y duración suficiente, incluso en formatos virtuales asincrónicos (Darling-Hammond et al., 2017). Bajo este marco, se ubica la estrategia propuesta, orientada a fortalecer creencias, conocimientos y prácticas docentes para promover y evaluar la argumentación científica en el aula universitaria.

En este sentido, la integración del PCK-A se enriquece al considerar la complejidad y las tensiones entre distintos marcos teóricos de la argumentación.

El enfoque sociodiscursivo resalta la construcción colectiva del conocimiento y la dimensión pragmática del diálogo (Plantin, 2009); el Patrón de Argumentación de Toulmin aporta una estructura analítica de componentes (Toulmin, 2007); por su parte, el modelo dialógico de Walton (1996) se centra en esquemas de interacción que facilitan la evaluación del progreso argumentativo. Estudios recientes evidencian que las creencias, conocimientos y prácticas del profesorado median la aplicación coherente de estos marcos y determinan la solidez de los argumentos en el aula universitaria (Wess et al., 2023).

Desde la perspectiva del PCK-A, reconocer las convergencias (la necesidad de evidencias explícitas, la refutación fundamentada y el diálogo crítico), así como sus divergencias metodológicas, permite diseñar tareas formativas más precisas y guiar la retroalimentación de manera alineada con la práctica científica. En síntesis, estos marcos teóricos permiten comprender cómo las creencias, conocimientos y prácticas docentes median la calidad de la argumentación en el aula universitaria y orientan la selección de criterios formativos para su promoción y evaluación.

En el marco de esta estrategia, se asume que la evaluación más adecuada de la argumentación docente combina instrumentos que captan creencias y prácticas con medidas de desempeño en tareas de análisis de argumentos (Sampson y Clark, 2006a; Sandoval y Millwood, 2007). En esa línea, el CUEFOAR operativiza dos dimensiones: ARES, desde el PCK-A, aproxima el dominio conceptual y epistémico; FOAR, el saber didáctico y de gestión del aula (Osborne et al., 2013). Como complemento, el AT, anclado en el TAP de Toulmin (2007) y en principios dialógicos de Walton (1996), ofrece una medida directa del uso de evidencias, garantías y refutaciones. La lectura integrada de CUEFOAR-AT permite detectar relaciones entre lo que el profesorado declara y lo que reconoce en argumentos reales, así como divergencias diagnósticas que orientan acciones formativas (Erduran y Jiménez-Aleixandre, 2007; Osborne et al., 2004).

Método

Este estudio se enmarca en un diseño cuasiexperimental de tipo pretest-postest sin grupo control, de carácter piloto y exploratorio, con muestreo no probabilístico por conveniencia. Participaron cinco docentes universitarios (D1-D5) de ciencias naturales (tres de física, uno de química y uno de biología), cuyas características demográficas, formación académica y experiencia profesional se presentan en la Tabla 1. En conjunto, las edades del grupo oscilan entre 29 y 36 años. Cuentan con formación de maestría o doctorado y experiencia profesional universitaria entre 0 y 10 años. El tamaño de la muestra ($n = 5$) responde al carácter piloto e instrumental del estudio, cuyo propósito fue explorar tendencias iniciales en el impacto de la estrategia y refinar los instrumentos antes de su aplicación a escala mayor.

Tabla 1*Descripción de las personas participantes*

Código de identificación	Género	Edad	Experiencia profesional	Experiencia profesional universitaria	Nivel de formación	Área	Nivel en el que enseña
D1	M	34	11-20 años	6-10 años	Doctorado	Química	Posgrado
D2	M	36	11-20 años	6-10 años	Maestría	Física	Pregrado
Aprovechamiento tecnológico	M	33	6-10 años	0-6 años	Maestría	Física	Pregrado
Desempeño docente	M	30	0-6 años	0-6 años	Maestría	Física	Pregrado
Interacción intercultural	F	29	6-10 años	0-6 años	Maestría	Biología	Posgrado

La intervención se desarrolló en modalidad virtual asincrónica mediante tres módulos que abordaron fundamentos de la argumentación científica, estrategias para promoverla en el aula universitaria y criterios para su evaluación. Cada módulo combinó lecturas guiadas, análisis de casos y elaboración de productos aplicados a la práctica docente. Asimismo, la dedicación estimada fue flexible, dentro de un rango aproximado de 8 a 12 horas según el ritmo de cada participante (Tabla 2). El proceso incluyó acompañamiento docente mediante retroalimentación escrita a cada producto entregado.

Tabla 2*Estructura general de la estrategia FAEUC*

Módulo	Objetivo	Contenidos	Actividades	Duración	Recursos
Módulo 1: Fundamentos de la argumentación científica	Proporcionar a los docentes una comprensión sólida de los conceptos clave en torno a la argumentación científica y su importancia en el aprendizaje de las ciencias.	Introducción a la argumentación científica	Lectura y reflexión	1 h	Guía y lectura
		Estructura de la argumentación científica	Lectura y análisis	1 h	Guía y lectura
		La argumentación en el contexto universitario	Lectura y ensayo reflexivo	2 h	Guía y lectura
Producto	Ensayo reflexivo integrador que analice el impacto de la argumentación en su práctica docente actual (1500 – 2000 palabras).				
Módulo 2: Estrategias para promover la argumentación en el aula universitaria	Desarrollar estrategias efectivas que los docentes puedan aplicar en el aula para promover la construcción de argumentos sólidos y basados en evidencias.	Metodologías para promover la argumentación	Lectura y diseño de actividades	1 h	Guía y lectura

Módulo	Objetivo	Contenidos	Actividades	Duración	Recursos
Módulo 2: Estrategias para promover la argumentación en el aula universitaria	Desarrollar estrategias efectivas que los docentes puedan aplicar en el aula para promover la construcción de argumentos sólidos y basados en evidencias.	Estrategias dialógicas y diseño de actividades	Lectura y diseño de actividades	1 h	Guía y lectura
		Promoción de la argumentación colaborativa y diseño de actividades	Lectura y diseño de actividades	2 h	Guía y lectura
Producto	Ensayo reflexivo integrador que analice el impacto de la argumentación en su práctica docente actual (1500 – 2000 palabras).				
Módulo 3: Evaluación de la argumentación científica	Dotar a los docentes de herramientas para evaluar las habilidades argumentativas científicas de los estudiantes de manera eficaz y objetiva.	Criterios de evaluación de la argumentación Retroalimentación constructiva	Lectura	1 h	Lectura
		Aplicación de instrumento PCK argumentación	Desarrollo de instrumento	1 h	Instrumento
Producto	Desarrollo del instrumento para evaluar el Conocimiento Pedagógico del Contenido (PCK) en Argumentación Científica (McNeill et al., 2016).				

La modalidad asincrónica generó variaciones en los tiempos de avance: mientras algunos docentes completaron la secuencia en períodos breves, otros lo hicieron de manera más extendida, lo que permitió adaptar la participación a sus cargas laborales. La estrategia se implementó en febrero de 2025, tras la aplicación del pretest en octubre de 2024; el posttest se aplicó en julio de 2025. En total, la intervención tuvo una duración aproximada de seis meses, mediante actividades virtuales. Así, algunos participantes completaron la secuencia en una semana, mientras que otros la desarrollaron en períodos más extensos: D1 y D2 en aproximadamente cuatro meses, D3 y D4 en cerca de seis meses y D5 la finalizó en el lapso mínimo.

Instrumentos y procedimiento

El instrumento, denominado CUEFOAR y elaborado por el equipo investigador: (Martínez-Morales et al., 2026), consta de 22 ítems organizados en dos dimensiones: ARES (Argumentación científica y su estructura), con 14 ítems que exploran las creencias sobre la estructura y el uso de la argumentación científica dentro y fuera del aula; y FOAR (Fomento de la argumentación y desarrollo de habilidades argumentativas), con 8 ítems que indagan las concepciones y acciones docentes para promoverla. Cada dimensión se evalúa en tres niveles de complejidad (Básico, Medio, Superior), establecidos mediante consenso de expertos y análisis de dificultad de los ítems.

En tal sentido, el nivel Básico refleja reconocimiento de la importancia y componentes esenciales de la argumentación; el Medio, su integración en la planificación y en las actividades habituales de aula; y el Superior, la aplicación reflexiva y la evaluación

sistemática de argumentos. Las respuestas se registraron en una escala Likert de 1 a 5 que se transforma de manera lineal a una escala de 0 a 100 para permitir la comparación por categoría y por docente. El instrumento cuenta con evidencia previa de validez de contenido y estructura, así como consistencia interna adecuada (ver apartado de Fiabilidad y Validez).

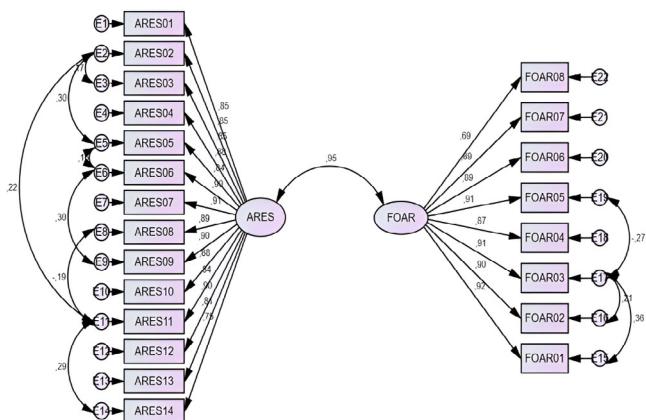
El segundo instrumento fue una traducción al español del *Argument Test* (AT) de Sampson y Clark (2006b), sin modificaciones en su estructura ni contenido. Se mantuvo la versión original debido a que evalúa habilidades de análisis lógico-argumentativo independientes del contexto cultural, además de que sus ejemplos no contienen referencias situadas que requieran adaptación. El AT consta de seis ítems organizados en dos partes: parte I, clasificación de argumentos por solidez; parte II, clasificación de contraargumentos por fuerza. La revisión asigna 1 punto por acierto y, adicionalmente, se analiza la desviación respecto del orden correcto de cada serie. Ambos instrumentos se incluyen como anexos del artículo —CUEFOAR (Anexo 1) y AT (Anexo 2)—.

Fiabilidad y validez

El CUEFOAR cuenta con evidencia sólida de fiabilidad y validez. Su validez de contenido fue establecida mediante juicio de siete expertos, con valores de V de Aiken entre 0.74 y 0.96. A nivel empírico, se aplicó a una muestra independiente de 150 docentes de ciencias, con lo que se obtuvo alta consistencia interna ($\alpha = .983$) y medida de muestreo adecuada ($KMO = .974$). El análisis factorial exploratorio identificó dos factores que explican el 77.7 % de la varianza y el análisis confirmatorio mostró índices de ajuste satisfactorios ($CFI = .984$; $TLI = .981$; $RMSEA = .047$) (ver Figura 1), lo que respalda la estabilidad de la estructura bidimensional ARES–FOAR. En el caso del AT, su uso se justificó por tratarse de una prueba de clasificación lógica basada en el TAP y en principios dialógicos universales, con validez de constructo ampliamente respaldada en estudios previos.

Figura 1

Modelo realizado con AMOS para el análisis factorial confirmatorio



Fuente: elaboración propia.

Análisis de datos

Se realizó un análisis descriptivo de los puntajes normalizados por docente y dimensión (media, DE, mínimo–máximo). Para el CUEFOAR, tras verificar la normalidad de las diferencias con la prueba de Shapiro–Wilk ($p > .05$), además de revisar la independencia de las observaciones y la ausencia de valores atípicos extremos, se aplicó la prueba t de Student para muestras relacionadas ($\alpha = .05$), a fin de comparar las mediciones pre y post; asimismo, se calculó el tamaño del efecto Cohen (d) para valorar la magnitud de los cambios. Para examinar la relación entre ARES–FOAR en la medición posterior, se utilizó la correlación de Pearson. En el caso del AT, se describieron los cambios pre–post en las medias, desviaciones estándar y en el error de ordenamiento. Finalmente, se realizó una triangulación CUEFOAR–AT a nivel de docente y por perfil de formación/experiencia para identificar convergencias y áreas de mejora. Todo el procesamiento y los análisis estadísticos se realizaron en el software *IBM SPSS Statistics* (versión 28).

Resultados

Creencias y prácticas de argumentación científica

Los resultados del CUEFOAR, que evalúa creencias sobre la estructura de la argumentación científica (ARES) y prácticas de fomento de la argumentación (FOAR) en tres niveles de complejidad (Básico [B], Medio [M], Superior [S]), muestran un incremento global en ambas dimensiones, con una tendencia positiva más marcada en FOAR que en ARES (ver Tabla 3 y Tabla 4). En ambos casos, la mejora se distribuye de forma desigual por subniveles. En ARES, el crecimiento más notorio aparece en el nivel Medio (+4 puntos), que agrupa afirmaciones relativas al uso de herramientas de simulación y visualización, la colaboración y el análisis crítico.

Esto sugiere que el componente didáctico–procedimental de la estructura argumentativa resultó especialmente sensible a la intervención, pues mostró la variación más pronunciada dentro de los subniveles, lo cual coincide con las actividades del Módulo 2 centradas en diseño didáctico y apoyos para el razonamiento con evidencias. En FOAR, la mejora más clara se observa en el nivel Superior (+7 puntos), vinculado al uso de estrategias activas (experimentos, debates, proyectos), creación de ambientes de participación y retroalimentación constructiva, lo que coincide con el enfoque de la estrategia formativa.

Tabla 3

Resultados CUEFOAR – PRE por docente y estadísticas descriptivas

Docente	ARES total	ARES S	ARES M	ARES B	FOAR total	FOAR S	FOAR M	FOAR B
D1	89,7	82,67	84	80	84,4	85	90	70
D2	80,7	68	80	100	90	95	90	60
D3	91,7	82,67	88	85	81,1	75	90	100

Docente	ARES total	ARES S	ARES M	ARES B	FOAR total	FOAR S	FOAR M	FOAR B
D4	100	86,67	100	100	100	100	100	100
D5	86,9	73,33	92	85	86,7	85	90	90
\bar{x}	89,8	78,66	88,8	90	88,4	88	92	84
SD	7,5	7,72	7,69	9,35	7,23	9,74	4,47	18,17
MIN	80,07	68	80	80	81,1	75	90	60
MAX	100	100	100	100	100	100	100	100

Tabla 4

Resultados CUEFOAR – POST por docente y estadísticas descriptivas

Docente	ARES total	ARES S	ARES M	ARES B	FOAR total	FOAR S	FOAR M	FOAR B
D1	87,6	77,33	92	75	91,1	95	90	70
D2	92,4	81,33	92	95	97,8	100	100	80
D3	100	86,67	96	100	94,4	95	90	100
D4	98,6	86,67	96	100	94,4	95	90	100
D5	82,1	72	84	85	82,2	85	80	70
\bar{x}	92,14	80,8	92,8	91	93,1	95	92	84
SD	7,5	6,3	5,93	10,83	6,96	6,12	8,36	15,16
MIN	82,1	72	84	85	82,2	85	80	70
MAX	100	100	100	100	100	100	100	100

Al mismo tiempo, los niveles Básicos de ARES y FOAR conservan medias más bajas y mayor dispersión, un patrón esperable en procesos donde la autoconciencia de límites, la evaluación crítica sistemática y la participación argumentativa respetuosa requieren de más tiempo, modelado y práctica deliberada para consolidarse. En la comparación estadística pre–post (Tabla 5), ninguna diferencia alcanzó significancia estadística ($\alpha = 0.05$), un resultado esperable dado el n reducido ($n = 5$); no obstante, los tamaños del efecto (d entre 0.13 y 0.63) resultan más informativos, con efectos pequeños a moderados que reflejan cambios relevantes en algunos docentes.

Aun así, las tendencias aportan información: FOAR Total aumenta 4,7 puntos y ARES–Medio, 4 puntos, lo que señala una mejora relevante en los terrenos donde la estrategia puso mayor énfasis (diseño didáctico, apoyos y evaluación de la argumentación). De forma complementaria, la correlación post entre ARES y FOAR es moderada ($r = 0,611$), lo que sugiere que un mayor dominio conceptual se asocia con mayor capacidad de promover prácticas argumentativas en el aula, aunque el tránsito del saber al hacer no sea lineal. Además de la prueba t , se calculó el tamaño del efecto de Cohen (d) para muestras relacionadas. Los valores oscilaron entre 0.13 y 0.63, lo que indica efectos de pequeños a moderados, en concordancia con el reducido tamaño muestral ($n = 5$).

Tabla 5
Comparación PRE-POST CUEFOAR

Dimensión	Media PRE	Media POST	Diferencia media	t-valor	p-valor	Cohen (d)
ARES Total	90	92	2,3	0,7	0,5	0,31
ARES S	79	81	2,1	0,7	0,5	0,31
ARES M	89	93	4	1	0,4	0,45
ARES B	90	91	1	0,3	0,8	0,13
FOAR Total	88	93	4,7	1	0,4	0,45
FOAR S	88	95	7	1,4	0,2	0,63
FOAR M	92	92	0	0	1	0,00
FOAR B	84	84	0	0	1	0,00

Argumentos y contraargumentos

En el AT (Tabla 6), la mejora es contundente. En la parte I (clasificación de argumentos por solidez), la media pasa de 1,8 a 4,8; en la parte II (clasificación de contraargumentos por fuerza), de 2,0 a 8,0; y el total, de 3,8 a 12,8. El patrón indica que, después de la estrategia, los docentes reconocen con mayor finura qué hace fuerte un argumento (uso de datos empíricos pertinentes, garantías explícitas que conectan datos con conclusión y respaldo que legitima dichas garantías) y qué hace sólida una objeción (evidencias contrarias relevantes, cuestionamiento al alcance de la garantía o refutaciones basadas en principios científicos o evidencia específica). Aunque la desviación estándar aumenta en el post, este comportamiento es coherente con escenarios donde algunos participantes pegan saltos muy grandes (p. ej., D1 de 3 a 18 en total), mientras otros progresan de forma más gradual (p. ej., D2 de 5 a 9).

Tabla 6

Respuestas y promedios de las respuestas de los docentes en el AT (pre y post)

D	Puntaje Parte I Pre	Puntaje Parte I Post	Puntaje Parte II Pre	Puntaje Parte II Post	Puntaje total Pre	Puntaje total Post
D1	1	8	2	10	3	18
D2	3	1	2	8	5	9
D3	1	6	1	5	2	11
D4	2	2	4	9	6	11
D5	2	7	1	8	3	15
\bar{x}	1,8	4,8	2	8	3,8	12,8
SD	0,75	2,79	1,10	1,67	1,47	3,25

El error de ordenamiento corresponde a la distancia entre el orden correcto de los argumentos y el orden asignado por cada docente; valores menores indican

una comprensión más fina de la solidez relativa entre opciones. El análisis de ordenamiento (Tabla 7) complementa los resultados generales, el error promedio disminuye de ~2,31 a ~1,03, con descensos en cada participante. Este resultado es especialmente sensible a la alineación con criterios de calidad de la evidencia, pues exige no solo identificar respuestas correctas, sino también ordenar opciones por su grado de solidez, un indicador de comprensión estructural de la calidad argumentativa.

Tabla 7

Promedios diferencias en el orden de argumentos AT (pre y post)

D	Pre				Post				Pre + Post	
	\bar{x} P I	\bar{x} P II	\bar{x} T	SD	\bar{x} P I	\bar{x} P II	\bar{x} T	SD	\bar{x}	SD
D1	2,44	2,44	2,44	0	0,78	0,67	0,72	0,04	1,28	0,86
D2	2,5	2,17	2,33	0,13	1,56	0,78	1,17	0,32	1,75	0,63
D3	2,67	2,61	2,64	0,02	1,22	1,67	1,44	0,18	2,04	0,61
D4	2,06	1,61	1,83	0,18	1	0,89	0,94	0,04	1,39	0,46
D5	2,39	2,17	2,28	0,09	0,89	0,89	0,89	0	1,56	0,70
\bar{x}	2,41	2,2	2,31	0,08	1,09	0,98	1,03	0,12	1,67	0,65
SD	0,2	0,34	0,27	0,06	0,28	0,35	0,25	0,12	0,27	0,13

El análisis por ítem (Tabla 8 y Tabla 9) confirma un cambio de criterio. Los ítems que concentraban tasas altas de error en el pre (p. ej., 1.1, 1.3, 1.5, 1.6 en la parte I; 4.1, 5.4 y 6.2 en la parte II) presentan aumentos notables de acierto en el postest, con menor dispersión en varios casos. Persisten, sin embargo, focos más resistentes (p. ej., 1.6, 5.2, 6.2), típicamente aquellos que exigen distinguir con precisión entre evidencia empírica directa y respaldos menos robustos (autoridad, analogía o principios generales no anclados en datos del caso).

Tabla 8

Porcentaje de las respuestas de los docentes en la primera parte del AT (pre y post)

Ítem	Pre								Post							
	C	%	ln1	%	ln2	%	ln3	%	C%	ln1	%	ln2	%	ln3	%	
1.1	3	0	4	60	2	40			20	4	60	5	20			
1.2	6	0	1	40	2	20	3	20	0	1	40	5	40	4	20	
1.3	2	0	3	60	4	20	5	20	20	1	60	5	20			
1.4	4	20	6	60	2	20			20	3	80					
1.5	5	20	1	40	2	20			20	6	80					
1.6	1	0	6	40	2	20	3	20	0	2	80	6	20			
2.1	3	20	2	60	5	20			20	4	40	5	20	6	20	
2.2	2	20	5	60	6	20			40	1	60					

Ítem	Pre								Post						
	C	%	ln1	%	ln2	%	ln3	%	C%	ln1	%	ln2	%	ln3	%
2.3	5	0	4	60	3	40			20	4	40	6	40		
2.4	4	20	6	40	1	20			20	3	60	2	20	4	20
2.5	6	0	1	80	3	20			40	5	60				
2.6	1	20	6	80					40	2	40	3	20		
3.1	6	0	3	80	1	20			60	4	20	5	20		
3.2	2	0	5	80	4	20			60	1	40				
3.3	5	0	1	40	2	40	4	20	60	3	20	6	20	6	20
3.4	1	0	2	40	3	20	5	20	20	2	40	3	40		
3.5	4	60	4	60	1	20	2	20	20	1	40	3	40		
3.6	3	0	1	60	2	20	6	20	0	4	60	5	20	6	20

Nota. C = Correcta; ln1/ln2/ln3 = Posiciones incorrectas por grado de solidez.

Tabla 9

Porcentaje de las respuestas de los docentes en la segunda parte del AT (pre y post)

Ítem	Pre								Post						
	C	%	ln1	%	ln2	%	ln3	%	C%	ln1	%	ln2	%	ln3	%
4.1	1	0	5	60	4	40			20	3	40	4	40		
4.2	5	0	4	40	1	20	2	20	60	4	40				
4.3	2	40	3	40	5	20			40	1	40	5	20		
4.4	3	0	6	80	1	20			0	2	60	1	40		
4.5	6	0	3	40	2	20	4	20	80	5	20				
4.6	4	0	6	80	1	20			20	3	60	6	20		
5.1	6	0	3	40	5	40	4	20	80	1	20				
5.2	4	0	3	40	1	20	2	20	20	3	40	5	40		
5.3	5	0	2	40	4	40	3	20	0	2	60	1	20	3	20
5.4	3	0	1	60	2	20	5	20	0	4	60	2	40		
5.5	1	0	5	40	2	20	3	20	80	5	20				
5.6	2	40	3	40	4	20			0	3	40	4	20	5	20
6.1	2	20	5	80	1	0			40	3	40	4	20		
6.2	5	20	2	40	3	20	4	20	60	3	20	4	20		
6.3	3	0	6	80	1	20			40	2	40	5	20		
6.4	6	20	1	80	2	0			100						
6.5	1	0	4	60	3	40			100						
6.6	4	60	3	40	1	0			60	2	20	5	20		

Nota. C = Correcta; ln1/ln2/ln3 = Posiciones incorrectas por grado de solidez.

Discusión de los resultados

La persistencia de la dificultad para responder de forma correcta ciertos ítems (en particular, 1.6, 5.2 y 6.2), que exigen distinguir evidencia empírica directa de

respaldos débiles o evaluar la pertinencia de las garantías frente a los datos del caso, es consistente con la literatura. El tránsito desde la retórica de las buenas razones hacia la justificación empírica y el abordaje de la refutación con estándares claros exige práctica contextualizada, ejemplos de contraste y retroalimentación por componente del argumento (Toulmin, 2007; Erduran et al., 2004; Zohar y Nemet, 2002). Asimismo, la triangulación CUEFOAR-AT por docente permite entender cómo se conectan las creencias y prácticas autoinformadas con el desempeño analítico en tareas de clasificación. Hay casos de convergencia clara, como D1 y D3, donde el alto FOAR post y el alto ARES post conviven con saltos contundentes en AT, lo que sugiere que el fortalecimiento del apoyo didáctico y de las concepciones estructurales de la argumentación se traduce en criterios más finos para valorar evidencias, garantías y alcance de las conclusiones.

La triangulación CUEFOAR-AT muestra patrones diferenciados: algunos docentes presentan convergencia clara entre creencias/prácticas y desempeño analítico, mientras que otros evidencian brechas entre disposición didáctica y criterios epistémicos para jerarquizar argumentos. Estas divergencias orientan los énfasis formativos (trabajo explícito con garantías, respaldos y contraste de evidencias).

Los resultados de la aplicación pre y post muestran un panorama consistente del efecto de la estrategia formativa sobre las creencias, los conocimientos y las prácticas de argumentación científica del profesorado, así como sobre su desempeño al jerarquizar argumentos y objeciones. En primer lugar, los puntajes del CUEFOAR evidencian incrementos en ambas dimensiones, con patrones diferenciados por nivel (básico, medio y superior) y variabilidad entre docentes. En segundo lugar, el AT arroja avances tanto en la parte I (clasificación de argumentos) como en la parte II (cuestionamientos o refutaciones), junto con una reducción del error de ordenamiento; esto sugiere una mayor precisión en los criterios epistémicos luego de la intervención.

Por último, la triangulación CUEFOAR-AT por docente identifica convergencias, cuando el reforzamiento de creencias/prácticas se acompaña de mejor desempeño analítico, y divergencias diagnósticas, cuando la disposición y el ambiente de aula aún no se traducen plenamente en criterios finos para distinguir la solidez de las cadenas evidencia-garantía-conclusión, lo que aporta elementos concretos para ajustar los énfasis formativos.

El contraste de los hallazgos con la teoría refuerza tres ejes complementarios. En primer lugar, el TAP continúa siendo una sintaxis operativa para valorar la calidad epistémica del argumento: la mejora en el ordenamiento del AT y en ARES-Medio indica un avance en la explicitación de datos, garantías, respaldos y refutaciones, coherente con lo reportado por Toulmin (2007), Erduran et al. (2004) y Kelly y Takao (2002). En segundo lugar, los marcos dialógicos muestran su efecto en la

regulación de la interacción: el aumento en FOAR-Superior y la reducción del error de ordenamiento reflejan progresos en normas como carga de la prueba, pertinencia y respuesta a objeciones, en línea con Walton (1996) y con la evidencia sobre comunidades de indagación (Osborne et al., 2004; Osborne et al., 2013). En tercer lugar, las divergencias entre FOAR y AT en algunos perfiles confirman lo planteado por el PCK-A: el tránsito del saber al hacer requiere decisiones didácticas específicas —selección de tareas ricas en evidencia, uso de rúbricas por componente TAP, modelación de garantías y refutaciones y retroalimentación formativa—, tal como señalan Sandoval et al. (2019), Osborne et al. (2013), Zohar (2007) y Zohar y Resnick (2021). Estos tres marcos, articulados, explican por qué ciertos elementos avanzan con mayor rapidez mientras otros requieren más tiempo, modelado y práctica deliberada para estabilizarse.

El tipo de desarrollo profesional ofrecido —breve, modular, centrado en problemas, con materiales y tareas específicas— se alinea con las recomendaciones para un desarrollo profesional docente eficaz: pertinencia del contenido, oportunidades de práctica, retroalimentación de calidad y duración suficiente para sostener el cambio (Darling-Hammond et al., 2017). La mejora en la parte II de AT es congruente con enfoques de indagación basada en argumentos y con la literatura que reporta mejoras cuando se privilegian tareas con datos y discusión de refutaciones (Sampson y Blanchard, 2012). Asimismo, el peso de la retroalimentación por componente del argumento, frecuente en la estrategia, coincide con la evidencia de su efecto sobre el aprendizaje (Hattie y Timperley, 2007).

Hallazgos afines reportan incrementos en la identificación de estructuras de Toulmin luego de intervenciones focalizadas (Çoban et al., 2016), a la vez que subrayan que los elementos más complejos: en particular, refutación y respaldo son los que más tiempo y exemplificación requieren para estabilizarse (Osborne et al., 2004). Finalmente, la idea de integrar herramientas digitales como apoyo para hacer visible la relación evidencia–conclusión y para promover la colaboración coincide con estudios sobre conocimiento tecnológico pedagógico del contenido (Chai et al., 2013), aunque también se reconoce la importancia de la mediación docente.

De igual manera, la lectura por docente aporta matices a estos resultados. Perfiles con formación avanzada y mayor experiencia, como D1 (Doctorado, 11–20 años de experiencia y 6–10 en universidad), tienden a mostrar progresos más pronunciados en la articulación entre apoyo didáctico y las pautas de justificación, lo que se traduce en mejoras sustantivas en AT y consolidación en FOAR. En trayectorias emergentes (D3, D4, D5), los incrementos en el AT se observan cuando las tareas ponen énfasis en el uso de datos propios, en estándares explícitos para justificar las afirmaciones y en refutaciones con criterios. Sin embargo, pueden mantenerse variaciones en ARES si no se sostiene la práctica reflexiva sobre la estructura del argumento y la guía

explícita del proceso de razonamiento. El caso de D2 ilustra esta desconexión: FOAR alto, ARES al alza, pero caída en la parte I de AT. Aquí, la ruta de mejora incluye ejercicios focalizados de descomposición de argumentos reales y comparación entre argumentos fuertes y débiles. Aun con un número de participantes reducido, la evidencia apunta a que la experiencia y la formación avanzada pueden facilitar la transferencia del conocimiento sobre argumentación a decisiones didácticas, mientras que los perfiles en consolidación requieren un modelado intensivo, parámetros específicos para la evaluación de los argumentos y procesos de práctica y retroalimentación para afianzar dichos elementos.

En tal sentido, los cambios observados se asocian a componentes específicos del DP: el trabajo con tareas basadas en evidencia incidió en la mejora del AT; la retroalimentación por componente contribuyó al aumento en FOAR-Superior; y la modelación de garantías y refutaciones favoreció los avances en ARES-Medio. Cabe destacar que estos resultados deben interpretarse considerando posibles sesgos de autoinforme, especialmente en FOAR, donde la percepción de promover prácticas argumentativas puede sobreestimarse en comparación con el desempeño analítico observado en el AT.

Implicaciones para la formación docente

De estos hallazgos se desprenden implicaciones para la formación docente y para líneas futuras de investigación. En el plano didáctico, resulta decisivo explicitar el TAP y usar rúbricas por componente con ejemplos de contraste, de modo que el estándar quede visible para docentes y estudiantes (Toulmin, 2007; Felton et al., 2022; Li et al., 2022; Erduran y Jiménez-Aleixandre, 2007). Los hallazgos sugieren la utilidad de normas dialógicas simples: si se afirma, se aportan datos; si se cuestiona, se ataca la garantía o se presentan datos de contraste; si se refuta, se acota el alcance, con el fin de mejorar el nivel de fundamentación y control de evidencias en el debate y de privilegiar datos y razones sobre la fuente (Giri y Paily, 2020; Walton, 1996; Osborne et al., 2004).

En términos de personalización, los resultados sugieren adaptar los apoyos según el perfil docente. Quienes muestran FOAR alto, pero dificultad en la parte I del AT, se benefician de un trabajo focalizado con garantías y respaldos; mientras que aquellos con fundamentos conceptuales sólidos (ARES alto), pero FOAR en consolidación, requieren acompañamiento en el diseño de tareas, la gestión del diálogo y la evaluación entre pares. Estas necesidades diferenciadas se complementan con la secuenciación de actividades —como indagaciones ABI, debates con evidencia y estudios de caso— y con retroalimentación basada en criterios específicos, lo cual acelera la adquisición de habilidades argumentativas (Sampson y Blanchard, 2012; Hattie y Timperley, 2007).

En el plano organizacional, los hallazgos apuntan a la importancia de transitar de intervenciones puntuales a ciclos breves y recurrentes de desarrollo profesional, apoyados en comunidades de práctica que comparten guiones, bancos de evidencias y rúbricas comunes, y que permitan sostener el avance en los componentes del TAP y la calidad del diálogo. En el plano de la investigación, se abren rutas para fortalecer la comparación entre fuentes y métodos, y para verificar la estabilidad de los resultados en situaciones reales de aula. Un primer eje consiste en incorporar observaciones de aula con protocolos de codificación basados en TAP, así como análisis de productos estudiantiles, de manera que la convergencia CUEFOAR-AT se contraste con evidencias de práctica efectiva. Un segundo eje contempla estudiar efectos diferenciales por disciplina (física, química, biología) y por nivel (pregrado/posgrado), para indagar si ciertas áreas requieren apoyos específicos en tipos de evidencia o en estilos de refutación. Un tercer eje implicaría evaluar la retención a mediano plazo y la transferencia a contextos nuevos.

Asimismo, el vínculo con tecnologías didácticas puede examinarse con instrumentos que combinen medidas de PCK y desempeño en tareas, como sugieren los trabajos sobre conocimiento tecnológico pedagógico del contenido (Chai et al., 2013). En relación con las recomendaciones para el desarrollo profesional efectivo, se plantea desarrollar formatos híbridos —no solo asincrónicos— que potencien la retroalimentación oportuna y el aprendizaje entre pares (Darling-Hammond et al., 2017).

Conclusiones

La estrategia formativa orientada a fortalecer las competencias docentes para promover y evaluar el uso de la argumentación científica en el aula universitaria de ciencias favoreció un avance general en las dimensiones analizadas. En CUEFOAR se observaron incrementos consistentes en ambas dimensiones, especialmente en aspectos relacionados con el uso de herramientas, la colaboración y las estrategias activas. En el AT, se registraron mejoras claras en la jerarquización de argumentos y cuestionamientos, acompañadas de una disminución del error de ordenamiento. Además, la correlación moderada entre fundamentos (ARES) y prácticas (FOAR) en el post sugiere consistencia, mientras que las diferencias individuales observadas indican que consolidar criterios epistémicos fuertes requiere trabajo continuo con rúbricas por componente, normas dialógicas estables, tareas con datos y retroalimentación formativa.

Por sus limitaciones, estos hallazgos deben interpretarse con cautela al considerar su generalización. En primer lugar, el tamaño muestral reducido ($n = 5$) restringe el poder estadístico de las pruebas t y la precisión de la estimación de efectos; por ello, los resultados se presentan como tendencias exploratorias. Además, la ausencia de grupo control impide formular inferencias causales sólidas. Por otra parte, el

carácter autoinformado de CUEFOAR, aun con el AT como complemento, requiere triangulación mediante observaciones de aula y análisis de tareas específicas.

También se debe considerar que la adaptación del AT fue una traducción sin ajustes conceptuales; aunque conserva la estructura original, necesita verificaciones psicométricas futuras en el contexto local.

A su vez, la intervención tuvo una duración reducida; los cambios en niveles básicos suelen requerir procesos más prolongados y múltiples ciclos de práctica y retroalimentación. Asimismo, la modalidad asincrónica implica variabilidad en el nivel de acompañamiento y en la intensidad del trabajo reflexivo, lo que puede influir en la estabilidad y profundidad de los avances observados. Finalmente, aunque se documentan de forma transparente los patrones pre–post, su interpretación debe situarse en el marco exploratorio de una experiencia con alta densidad formativa, pero alcance muestral modesto.

En términos de recomendaciones para el desarrollo profesional, los datos indican que se deben desarrollar procesos modulares, centrados en tareas y retroalimentación, así como personalización por perfil. Para futuras investigaciones, la integración de observaciones de aula, análisis de productos estudiantiles, estudios por disciplina y mediciones de durabilidad del aprendizaje y de su transferencia permitirá afinar el diagnóstico y optimizar la estrategia. Aun con las limitaciones señaladas, las variaciones, junto con los patrones de la triangulación CUEFOAR–AT y la coherencia con el marco teórico expuesto, respaldan que la estrategia fortaleció, con impacto significativo y relevancia pedagógica, las competencias argumentativas del profesorado de ciencias y las condiciones de aula que favorecen su puesta en práctica. En conjunto, el estudio ofrece evidencia inicial sobre cómo una estrategia de desarrollo profesional mediada por tecnología puede articular creencias, conocimientos y desempeño analítico, al tiempo que aporta un marco útil para futuras intervenciones en argumentación científica universitaria.

Contribución de autoría:

Alejandra Martínez-Morales: Conceptualización, metodología, investigación, curación de datos, redacción – borrador original, visualización, administración del proyecto.

Milany Gómez Betancur: Análisis formal, validación, redacción – revisión y edición.

Financiamiento: Este estudio no recibió financiamiento de ningún tipo.

Conflicto de interés: Las autoras declaran no tener conflicto de interés.

Declaración de ética: El proyecto cuenta con aprobación del Comité de Ética de la Facultad de Educación de la Universidad Antonio Nariño.

Data set: <https://data.mendeley.com/datasets/9pbt3jdzcf/1>

Referencias bibliográficas

- Adúriz-Bravo, A. (2014). Revisiting school scientific argumentation from the perspective of the history and philosophy of science. En M. Matthews. (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 1443–1472). Springer Netherlands. https://doi.org/10.1007/978-94-007-7654-8_45
- Archila, P., Molina, J., Truscott, A.-M., & Restrepo, S. (2022). Drama as a powerful tool to enrich socioscientific argumentation. *International Journal of Science and Mathematics Education*, 21, 1661–1683. <https://doi.org/10.1007/s10763-022-10320-3>
- Benarroch, A., Briceño-Martínez, J.-J., & Bernal-Ballen, A. (2024). Analysis of conceptions and beliefs, argumentative teaching practices, and reflection of university teachers: A case study on teachers' professional development. *Sage Open*, 14(3). <https://doi.org/10.1177/21582440241281352>
- Chai, C., Hwee Ling Koh, J., & Tsai, C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society*, 16(2), 31–51. <https://www.jstor.org/stable/jeductechsoci.16.2.31>
- Çoban, G., Akpinar, E., Baran, B., Sağlam, M., Özcan, E., & Kahyaoğlu, Y. (2016). The evaluation of “Technological pedagogical content knowledge based argumentation practices” training for science teachers. *Egitim ve Bilim*, 41(188), 1–33. <https://doi.org/10.15390/EB.2016.6615>
- Darling-Hammond, L., Hyler, M., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. <https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report>
- Erduran, S., & Jiménez-Aleixandre, M. (2007). *Argumentation in Science Education: Perspectives from classroom-based research*. Springer, Dordrecht. <https://doi.org/10.1007/978-1-4020-6670-2>
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the use of Toulmin's Argument Pattern in studying science discourse. *Science Education*, 88(6), 915–933. <https://doi.org/10.1002/sce.20012>
- Faize, F., Husain, W., & Nisar, F. (2018). A critical review of scientific argumentation in science education. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(1), 475–483. <https://doi.org/10.29333/ejmste/5274>
- Felton, M., Levin, D., De La Paz, S., & Butler, C. (2022). Scientific argumentation and responsive teaching: Using dialog to teach science in three middle-school classrooms. *Science Education*, 106(6), 1354–1374. <https://doi.org/10.1002/sce.21740>
- Giri, V., & Paily, M. (2020). Effect of scientific argumentation on the development of critical thinking. *Science & Education*, 29(3), 673–690. <https://doi.org/10.1007/s11191-020-00120-y>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Jiménez-Aleixandre, M., & Erduran, S. (2007). Argumentation in science education: An overview. En S. Erduran, & M. Jiménez-Aleixandre. (Eds.), *Argumentation in Science Education: Perspectives from Classroom-Based Research* (pp. 3–27). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6670-2_1
- Kelly, G., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86(3), 314–342. <https://doi.org/10.1002/sce.10024>

- Li, X., Wang, W., & Li, Y. (2022). Systematically reviewing the potential of scientific argumentation to promote multidimensional conceptual change in science education. *International Journal of Science Education*, 44(7), 1165–1185. <https://doi.org/10.1080/09500693.2022.2070787>
- McNeill, K., González-Howard, M., Katsh-Singer, R., & Loper, S. (2016). Pedagogical content knowledge of argumentation: Using classroom contexts to assess high-quality PCK rather than pseudoargumentation. *Journal of Research in Science Teaching*, 53(2), 261–290. <https://doi.org/10.1002/tea.21252>
- Martínez-Morales, A. y Briceño-Martínez, J. (2025). Revisión sistemática sobre la argumentación en la formación del profesorado de Ciencias en Activo. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 22(3), 3603. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2025.v22.i3.3603
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50(3), 315–347. <https://doi.org/10.1002/tea.21073>
- Plantin, C. (1998). *La argumentación*. Editorial Ariel.
- Plantin, C. (2009). A place for figures of speech in argumentation theory. *Argumentation*, 23, 325–337. <https://doi.org/10.1007/s10503-009-9152-0>
- Sampson, V., & Blanchard, M. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, 49(9), 1122–1148. <https://doi.org/10.1002/tea.21037>
- Sampson, V., & Clark, D. (2006a). Assessment of argument in Science Education: A critical review of the literature. *ICLS 2006 Proceedings*, 655–661. <https://repository.isls.org//handle/1/3571>
- Sampson, V., & Clark, D. (2006b). The development and validation of the nature of science as argument questionnaire (NSAAQ). *Paper presented at the Annual Conference of the National Association for Research in Science Teaching*. San Francisco, CA.
- Sandoval, W., & Millwood, K. (2007). What can argumentation tell us about epistemology? En S. Erduran, & M. Jiménez-Aleixandre. (Eds.), *Argumentation in Science Education* (pp. 71–88). Springer. https://doi.org/10.1007/978-1-4020-6670-2_4
- Sandoval, W., Enyedy, N., Redman, E., & Xiao, S. (2019). Organizing a culture of argumentation in elementary science. *International Journal of Science Education*, 41(13), 1848–1869. <https://doi.org/10.1080/09500693.2019.1641856>
- Toulmin, S. (2007). *Los usos de la argumentación*. Editorial Península.
- Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Routledge. <https://doi.org/10.4324/9780203811160>
- Wess, R., Priemer, B., & Parchmann, I. (2023). Professional development programs to improve science teachers' skills in the facilitation of argumentation in science classroom—a systematic review. *Disciplinary and Interdisciplinary Science Education Research*, 5(9), 1–22. <https://doi.org/10.1186/s43031-023-00076-3>

- Zohar, A. (2007). Science teacher education and professional development in argumentation. In M. Jiménez-Aleixandre, & S. Erduran. (Eds.), *Argumentation in Science Education: Perspectives from Classroom-Based Research* (pp. 245–269). Springer. https://doi.org/10.1007/978-1-4020-6670-2_12
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62. <https://doi.org/10.1002/tea.10008>
- Zohar, A., & Resnick, M. (2021). Professional development for the support of teaching through inquiry. En J. Lee, & D. Chan. (Eds.), *International Handbook of Inquiry and Learning* (pp. 109–129). Routledge.