PH PRO HOMINUM
REVISTA DE CIENCIAS SOCIALES Y HUMANAS

Artículo original

Retroalimentación con IA generativa y reflexión pedagógica en la formación inicial docente: una mirada desde la investigación acción

Feedback Using Generative AI and Pedagogical Reflection in Initial Teacher Education: A Perspective from Action Research

https://doi.org/10.47606/ACVEN/PH0617

Mónica Judith Ramos-Cáceres[1*]
https://orcid.org/0009-0008-1454-1584
mramosc@ispa.edu.pe

Trinidad Rocio Morales-Málaga[1]
https://orcid.org/0009-0007-3429-8115
tmoralesm@ispa.edu.pe

Angelica Maria Reyes-Neciosup[1]
https://orcid.org/0009-0007-9118-7331
areyesn@ispa.edu.pe

Constantino Paricela-Huancollo[1]
https://orcid.org/0000-0002-3856-7315
cparicelah@ispa.edu.pe

Edith Angela Suca-Yanpasi[1]
https://orcid.org/0009-0006-9973-2667
esucay@ispa.edu.pe

Recibido: 14/09/2026 **Aceptado:** 13/11/2026

RESUMEN

La retroalimentación del formador durante la práctica preprofesional suele ser lenta y con escaso grado de personalización. Se presentó un ciclo corto de investigación-acción participativa cuyo propósito fue analizar cómo la retroalimentación estructurada con inteligencia artificial generativa (IAG) promueve la reflexión pedagógica y la reestructuración didáctica de estudiantes de formación inicial docente (FID). El diseño se basó en el enfoque cualitativo y el paradigma sociocrítico. Abarcó cuatro semanas académicas y dos ciclos de planificación, acción, observación y reflexión. La audiencia fueron 21 practicantes organizados en cuatro Grupos de Estudio-Trabajo. Las peticiones a la IAG se estructuraron con los marcos PARTS, CLEAR y con la Escalera de Retroalimentación de Wilson. Los datos proceden del historial de prompts, el diario praxeológico, la observación de aula, el grupo focal y las entrevistas, y se codificaron en ATLAS.ti 9 mediante procesos de codificación abierta y axial, y una integración selectiva. Los resultados muestran un tránsito del nivel descriptivo al analítico y crítico, un aumento de las peticiones estructuradas y de la verificación crítica de las salidas, y coocurrencias intensas entre el feedforward y el rediseño didáctico, así como entre la deliberación grupal y la reflexión crítica. Se concluye que la ruta evidenciada es aplicable siempre que la IAG funcione como un interlocutor sometido al juicio del formador y del grupo.

Palabras Clave: inteligencia artificial generativa, retroalimentación formativa, reflexión pedagógica, formación inicial docente, investigación-acción participativa

1. Escuela de Educación Superior Pedagógica Pública de Arequipa (EESPPA), Perú
* Autor de correspondencia: mramosc@ispa.edu.pe

ABSTRACT

Feedback from instructors during pre-professional practicum is often slow and lacks personalization. A short cycle of participatory action research was conducted to analyze how structured feedback using generative artificial intelligence (GAI) promotes pedagogical reflection and instructional restructuring among students in initial teacher education (ITE). The design was based on a qualitative approach and the sociocritical paradigm. It spanned four academic weeks and two cycles of planning, action, observation, and reflection. The participants were 21 student teachers organized into four Study-Work Groups. Requests to the GAI were structured using the PARTS and CLEAR frameworks, as well as Wilson's Feedback Ladder. Data were drawn from the prompt history, the praxeological journal, classroom observations, focus group discussions, and interviews, and were coded in ATLAS.ti 9 using open and axial coding processes, followed by selective integration. The results show a transition from the descriptive to the analytical and critical levels, an increase in structured requests and critical verification of outcomes, and strong co-occurrences between feedforward and instructional redesign, as well as between group deliberation and critical reflection. It is concluded that the path identified is applicable provided that the IAG functions as a partner subject to the judgment of the trainer and the group.

Keywords: generative artificial intelligence, formative feedback, pedagogical reflection, initial teacher education, participatory action research

INTRODUCCIÓN

La práctica preprofesional es el escenario donde el estudiante de formación inicial docente (FID) confronta la teoría con el aula. El provecho de esta dependerá de una retroalimentación formativa que tenga claridad con respecto hacia dónde se avanza y qué paso sigue (Hattie & Timperley, 2007). En las escuelas pedagógicas, la elevada proporción de practicantes por formador tiende a obstaculizar esa devolución y a volverla netamente descriptiva, de manera que numerosos planes de sesión reciben comentarios tardíos o centrados en los aspectos de forma.

La inteligencia artificial generativa (IAG) permite ofrecer una respuesta inmediata a en ese sentido. Steiss et al. (2024) compararon la calidad de la retroalimentación humana con la de ChatGPT sobre los textos de estudiantes y encontraron que cada fuente sobresale en relación con criterios distintos. Banihashem et al. (2024) contrastaron la retroalimentación entre los pares con la generada por IA y describieron diferencias en el contenido de las dos versiones. Dentro de la FID, la retroalimentación sobre la escritura académica asistida por IAG ya se ha explorado (Ossa & Willatt, 2023), y dentro de esos alcances Hofmann et al. (2025) examinaron si la retroalimentación generada por IA promovía las habilidades reflexivas de docentes en formación. Por su parte, Farillon-Labis (2025) indagó las reflexiones de futuros profesores de matemáticas mientras planificaban clases con IAG.

El marco CLEAR propone que el prompt sea conciso, lógico, explícito, adaptativo y reflexivo, por lo que la utilidad de esas devoluciones dependería de la calidad de la petición (Lo, 2023). Al respecto, Sandoval Hernández et al. (2024) estudiaron el uso del prompt de ChatGPT como un asistente en la educación. De los trabajos anteriores se desprende que una petición ambigua produce una respuesta genérica, y que una petición con rol, destinatario, criterios y formato orientaría una retroalimentación anclada en la planificación del practicante.

La revisión de Zhai et al. (2024) sobre la dependencia excesiva de los sistemas de diálogo con IA generó señales de alarma en cuanto a un posible deterioro de las capacidades cognitivas del estudiantado. Una lectura histórica y crítica de la IA en la educación invita a examinar los supuestos y vacíos (Williamson & Eynon, 2020), y la revisión de Celik et al. (2022) sobre la IA para los docentes recogió tanto sus promesas como sus desafíos. Por su lado, Hwang et al. (2020), delinearon aspectos como la visión, los retos existentes, así como los roles y las cuestiones sobre investigación de la IA en educación.

La revisión sistemática de Martínez-Garcés et al. (2026) reunió 16 estudios empíricos sobre el uso de la IAG por el profesorado universitario latinoamericano. El impacto de la IA en la formación docente del CUNORTE fue abordado por Ceballos Macías (2025) y López Tineo et al. (2025) se centraron en la implementaron asistentes virtuales con IAG para mejorar la enseñanza docente. La reflexión de los practicantes durante las experiencias de campo fue sistematizada con el marco del modelo ONION (García Lazaro & Reyes de Cózar, 2024). Los estudios sobre la intención de adopción de la IAG por futuros docentes (Liu et al., 2025) y sobre su alfabetización mediante la indagación colaborativa (Kelley & Wenzel, 2025) han ido allanando el terreno, con un foco declarado no recae sobre la calidad de la reflexión pedagógica. Según lo explorado, la evidencia hispanohablante disponible se concentra en los usos y las percepciones, existiendo menor atención con respecto a los procesos reflexivos de los practicantes.

El artículo presente se articuló en torno a tres ejes que integran en primer lugar la retroalimentación formativa y la vertiente prospectiva (feedforward), que convierte la devolución en un insumo para decidir la planificación siguiente. El segundo fue el de la reflexión pedagógica, entendida la misma como el tránsito desde la narración descriptiva hacia la explicación analítica y, de allí, hacia el cuestionamiento crítico de los supuestos que guian los procesos de enseñanza y aprendizaje. El eje final fue la mediación de la IAG como un interlocutor dialógico, cuyas respuestas se tratan como objeto de análisis y no como verdad absoluta. Los Grupos de Estudio-Trabajo (GET) conectaron esos ejes porque allí la devolución algorítmica individual se sometía a una deliberación colectiva y a contraste teórico, y porque la investigación-acción participativa sitúa a los propios practicantes como agentes de su mejora.

Entre las fuentes revisadas no se localizó un diseño que articulase el prompting estructurado, y tampoco la deliberación en grupos de estudio y los ciclos de investigación-acción dentro de una escuela pedagógica de práctica intensiva. No abundan igualmente los modelos de ciclo corto que permitan ensayar esa articulación antes de un estudio extenso. Ese vacío orientó la pregunta PICO. En estudiantes de séptimo y octavo semestre de la FID (P), ¿de qué manera la retroalimentación estructurada con IAG integrada en ciclos de investigación-acción de cuatro semanas (I), frente a la retroalimentación convencional del formador sin IAG (C), promueve la reflexión pedagógica y la reestructuración didáctica (O)? La comparación debe asumirse como el contraste analítico entre el plan inicial y el plan reestructurado, y no como un grupo control, dado el carácter participativo y sociocrítico del diseño.

El objetivo general fue analizar cómo la retroalimentación con IAG, estructurada con los marcos PARTS y CLEAR y con la Escalera de Wilson e integrada en ciclos de investigación-acción, promovía la reflexión pedagógica y la reestructuración didáctica de estudiantes de la FID.

METODOLOGÍA

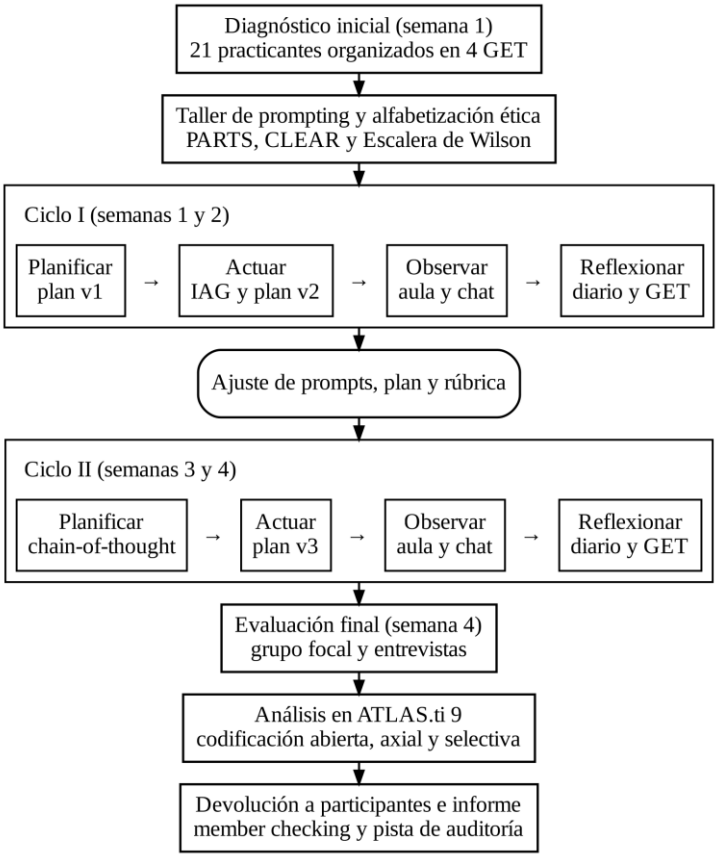
El estudio se planteó como una investigación-acción participativa, propia de un enfoque cualitativo y del paradigma sociocrítico. El diseño constó de dos ciclos dialécticos de dos semanas, cada uno con las fases de planificación, acción, observación y reflexión, articuladas con los momentos praxeológicos de ver, juzgar, actuar y devolver creativamente. La Figura 1 resume el protocolo completo implementado.

Participaron 21 practicantes de séptimo y octavo semestre de una escuela de educación superior pedagógica pública de Arequipa, distribuidos en cuatro Grupos de Estudio-Trabajo (GET) de cinco integrantes cada uno, con la excepción de uno de seis. El muestreo fue intencional. Los criterios de inclusión fueron la matrícula vigente, la práctica preprofesional en curso, la participación voluntaria y el consentimiento informado firmado. Los formadores de práctica acompañaron las sesiones y aplicaron la ficha de observación.

Tanto los practicantes y formadores firmaron un consentimiento informado antes del primer ciclo y conservaron el derecho de retiro sin tener el riesgo de consecuencias académicas. Los datos se anonimizaron con códigos (E-FID para estudiantes y F-DOC para formadores), y se acordó no ingresar a la IAG los nombres, imágenes o datos sensibles de los escolares. Cada plan de sesión debía declarar la trazabilidad de la interacción con la IAG para preservar la autoría pedagógica.

Figura 1

Protocolo de investigación-acción de cuatro semanas



Nota. Elaboración propia. GET= Grupo de Estudio-Trabajo. IAG= inteligencia artificial generativa.

Protocolo de prompting y plan de cuatro semanas

El taller inicial entrenó a los practicantes en tres recursos que fueron el marco PARTS, que organiza la petición según persona, objetivo, destinatarios, tema y estructura de la respuesta. Seguidamente se aclararon las pautas del enfoque CLEAR (Lo, 2023) que exige unas peticiones concisas, lógicas, explícitas, adaptativas y reflexivas. En tercer término, se explicó la Escalera de Retroalimentación de Wilson, que ordena la devolución en cuatro peldaños (saber, aclarar, valorar, plantear inquietudes y sugerir), y se alineó con las preguntas sobre la meta y el paso siguiente (Hattie & Timperley, 2007).

La plantilla de petición se validó con la revisión de dos formadores de práctica, quienes ajustaron la redacción de los criterios, y con un ensayo sobre los tres planes de sesión ajenos a la muestra. Por ejemplo, dentro del proceso, una petición difusa como "*Revisa mi clase de comunicación*" se reformuló como "*Actúa como un formador experto en didáctica de la comunicación y evalúa mi plan de sesión para los estudiantes de segundo grado de secundaria. Mi objetivo es aumentar la participación. Aclara primero mi propósito, valora lo logrado, plantea tus inquietudes y sugiere cambios en un máximo de 300 palabras*". La comparación entre ambas formas permitió al grupo anticipar la diferencia en torno a la calidad de las devoluciones de la IAG.

La semana inicial reunió el diagnóstico y el taller de prompting. La semana 2 correspondió a la primera versión del plan de sesión, implicando esto la revisión con la IAG, la ejecución de la versión reestructurada y la observación participativa del aula. En la tercera semana se profundizaron las peticiones de razonamiento paso a paso (chain-of-thought) y con la deliberación en el GET. La última semana (4) se dedicó a la evaluación final mediante un grupo focal y la realización de entrevistas semiestructuradas.

Fuentes de datos y análisis

El historial de prompts se exportó cada semana y los diarios fueron redactados tras cada sesión ejecutada, de modo que cada practicante aportó insumos para dos diarios. La ficha de observación se aplicó una vez por ciclo. El grupo focal se realizó en cada GET durante la semana 4, teniendo una duración aproximada de 60 minutos. Las entrevistas semiestructuradas fueron de unos 25 minutos y se aplicaron a seis practicantes y cuatro formadores.

Se emplearon cinco fuentes, donde se contemplaron la matriz de historial de prompts que registró la arquitectura de las peticiones y el diálogo con la IAG. Fue incorporado el diario praxeológico, que recogió el ver, el juzgar, el actuar y el devolver creativamente de cada practicante. La ficha de observación de aula fue aplicada por el formador. Los documentos del grupo focal y las entrevistas semiestructuradas se realizaron al cierre. La rúbrica de reflexión aplicada distinguió tres niveles que abarcaron desde la narración descriptiva hasta el cuestionamiento crítico de los supuestos.

La rúbrica de niveles de reflexión se definió con indicadores observables. El nivel descriptivo agrupó los diarios que narraban los hechos y ordenaron el tiempo de la sesión sin llegar a explicar las causas. El nivel analítico incorporó explicaciones apoyadas en la teoría o en la retroalimentación recibida. El nivel crítico cuestionó los supuestos propios, reconocía el peso del entorno escolar y proponía una reconstrucción de la práctica. Dos investigadores aplicaron la rúbrica de forma independiente a una muestra de diarios y discutieron las discrepancias hasta alcanzar consenso, procedimiento que sustenta la confirmabilidad del análisis y evitó los sesgos.

El análisis se realizó en ATLAS.ti 9, donde se creó un proyecto (unidad hermenéutica) con cinco grupos de documentos primarios, contándose con uno por fuente. La codificación abierta produjo 13 códigos. La fase axial los agrupó en siete familias y la

selectiva integró las dos categorías centrales: la IAG con la retroalimentación formativa y la reflexión pedagógica. El procedimiento siguió los principios del análisis temático reflexivo (Braun & Clarke, 2021). Las relaciones entre los códigos se estimaron con el coeficiente c de ATLAS.ti, calculado como c= n12 / (n1 + n2 - n12), donde n12 es el número de citas en que coocurren dos códigos y n1 y n2 son sus frecuencias. Para el rigor exigido se aplicaron la triangulación de fuentes, la triangulación de métodos, la comprobación con los participantes y una pista de auditoría con libro de códigos y memos.

RESULTADOS

El diagnóstico de la semana 1 (n= 21) mostró cuatro necesidades que fueron sintetizadas en la Tabla 1. De las peticiones iniciales diecisiete de ellas carecían de rol y de criterio, y producían devoluciones genéricas de la IAG. Dieciocho practicantes aceptaron las sugerencias sin atreverse a contrastarlas con la teoría ni con el currículo. En torno a los diarios, 14 se ubicaron en el nivel descriptivo, 6 en el analítico y solo uno en el crítico. Según la consulta, la retroalimentación del formador demoraba nueve días en promedio, siendo este un lapso que impedía su utilización antes de la sesión.

Tabla 1

Resumen de la fase de diagnóstico y acciones derivadas

Dimensión	Hallazgo	Necesidad identificada	Acción derivada
Formulación de prompts	17 de 21 peticiones iniciales sin rol ni criterio	Estructurar la petición	A1. Taller de prompting con PARTS, CLEAR y Escalera de Wilson
Verificación de la salida	18 de 21 aceptaron las sugerencias sin contrastarlas	Contrastar y verificar la salida de la IAG	A2. Protocolo de verificación crítica y ética
Nivel de reflexión	14 descriptivos, 6 analíticos y 1 crítico	Pasar de narrar a explicar y cuestionar	A3. Deliberación en el GET con rúbrica de niveles
Oportunidad de la devolución	Demora media de 9 días en la retroalimentación del formador	Contar con feedforward antes de la sesión	A1 y A3. Revisión con la IAG previa a la sesión y cierre en el GET

Nota. Elaboración propia. Las fuentes fueron la matriz de historial de prompts, los diarios praxeológicos y entrevistas breves.

Plan de acción y secuencia evaluativa

Las necesidades dieron origen a tres acciones estratégicas formativas. La primera fue un taller de prompting pedagógico con los ejes PARTS, CLEAR y la Escalera de Wilson. La siguiente consistió en un protocolo de verificación crítica y ética de las salidas de la IAG, lo que exigía contrastar cada sugerencia con un referente teórico o curricular. La tercera se instaló a partir de las sesiones de deliberación en el GET con la rúbrica de niveles de reflexión.

La secuencia evaluativa combinó desde una evaluación diagnóstica en la semana 1 hasta una evaluación sumativa en la semana 4. La Tabla 2 agrupa los resultados. Se logra apreciar que los diarios de nivel descriptivo descendieron de 14 a 3, y que los de nivel crítico ascendieron de 1 a 7. Por su parte, las peticiones estructuradas pasaron de 4 a 19 y la verificación crítica de 3 a 18. Los planes con rediseño sustancial llegaron a 17 en la semana 4, evidenciando todo esto ganancias importantes.

Tabla 2

Secuencia evaluativa y resultados por semana

Semana y tipo de evaluación	Nivel descriptivo	Nivel analítico	Nivel crítico	Prompts estructurados	Verificación crítica	Planes con rediseño sustancial
1. Diagnóstica	14	6	1	4	3	-
2. Formativa (ciclo I)	10	9	2	12	8	8
3. Formativa (ciclo II)	6	11	4	17	14	13
4. Sumativa	3	11	7	19	18	17

Nota. Frecuencias de practicantes por semana. Los tres niveles suman 21 en cada fila. Elaboración propia.

El avance fue desigual entre los GET porque los primeros diarios de nivel crítico aparecieron en dos GET durante la semana 2 y los restantes los evidenciaron en la semana 3. Dicha diferencia se atribuyó a la frecuencia de las reuniones de deliberación. Las entrevistas con los cuatro formadores (F-DOC-01 a F-DOC-04) describieron una menor carga de revisión escrita y una conversación más centrada en las decisiones didácticas. De estos diálogos surgió la necesidad de resguardar el tiempo para la retroalimentación socioemocional. En cuanto al uso ético, 17 practicantes llegaron a declarar la traza de la interacción con la IAG en el plan de la semana 4, todo esto en relación con un número de cinco en la semana 2.

Reestructuración didáctica de los planes de sesión

La matriz comparativa de los planes mostró que 17 de los 21 practicantes introdujeron en la semana 4 los cambios sustanciales con respecto a la versión inicial. Estos son entendidos como modificaciones de los objetivos, o también de la secuencia y de la evaluación. Un caso permitió ilustrar el patrón. La primera versión de un plan de comprensión lectora proponía el uso de las fichas de ejercicios aislados con un tiempo insuficiente. Tras la revisión con la IAG, la segunda versión diseñada recortó las explicaciones, la cual tras la deliberación en el GET y en una tercera intención, sustituyó la ficha por la elaboración de una carta de reclamo dirigida a una autoridad local, con la correspondiente rúbrica de coevaluación entre pares. El recorrido de la versión 1 a la 3 combinó un ajuste de la gestión del tiempo, el cambio del propósito comunicativo, así como una redefinición de la evaluación y todo esto articulado con el mayor protagonismo estudiantil. En tal sentido, se avanzó desde el nivel descriptivo hacia el crítico. Los cuatro practicantes sin rediseño sustancial conservaron el plan por la falta de tiempo de ejecución, según se desprende de los diarios.

Códigos y coocurrencias en ATLAS.ti 9

El proyecto de ATLAS.ti 9 reunió 119 documentos primarios. Se contemplaron 21 historiales de prompts, 42 diarios praxeológicos, 42 fichas de observación, cuatro transcripciones de acercamientos al grupo focal y diez de entrevistas. El libro de códigos agrupó 13 códigos en 7 familias y sumó 497 citas codificadas. La reflexión analítica (58 citas), el feedforward recibido (52) y el rediseño didáctico (51) mostraron el mayor enraizamiento. La alucinación detectada (12) fue el código menos frecuente.

La Tabla 3 presenta los coeficientes c del proyecto y la Figura 2 los representa como una red. Se observó que la coocurrencia más intensa vinculó el prompt estructurado con el feedforward recibido (c= 0,46), seguida esta de la deliberación en el GET con la reflexión

crítica (c= 0,44). El feedforward se asoció también con el rediseño didáctico (c= 0,39), igual que el diálogo iterativo lo hizo con la reflexión analítica (c = 0,39). La verificación crítica coocurrió con la reflexión analítica (c= 0,31) y con la detección de alucinaciones (c= 0,24). Por su parte, la reflexión descriptiva apenas coincidió con el prompt estructurado (c= 0,06).

Tabla 3

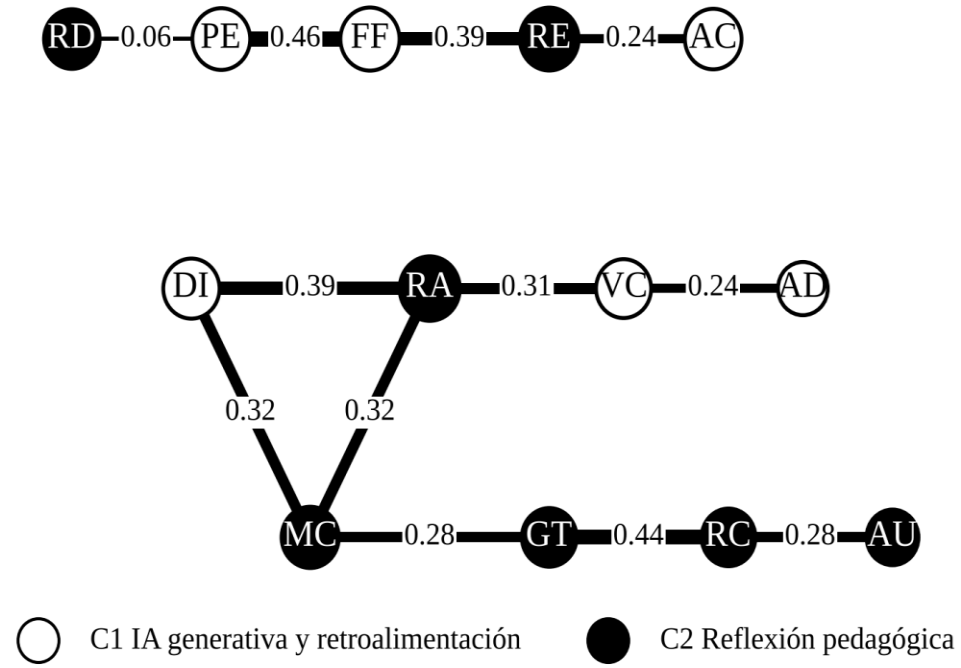
Coocurrencias entre los códigos en el proyecto de ATLAS.ti 9

Código 1	Código 2	Coocurrencias	Coeficiente c
Prompt estructurado	Feedforward recibido	31	0,46
Deliberación en el GET	Reflexión crítica	19	0,44
Feedforward recibido	Rediseño didáctico	29	0,39
Diálogo iterativo	Reflexión analítica	27	0,39
Metacognición	Reflexión analítica	25	0,32
Diálogo iterativo	Metacognición	20	0,32
Verificación crítica	Reflexión analítica	22	0,31
Reflexión crítica	Autonomía profesional	11	0,28
Deliberación en el GET	Metacognición	17	0,28
Andamiaje cognitivo	Rediseño didáctico	18	0,24
Verificación crítica	Alucinación detectada	9	0,24
Reflexión descriptiva	Prompt estructurado	5	0,06

Nota. El coeficiente c se calcula como n12 / (n1 + n2 - n12). Se muestran los 12 pares incluidos en la Figura 2.

Figura 2.

Red de coocurrencias de códigos



Nota. El grosor de la línea es proporcional al coeficiente c. Los nodos blancos corresponden a C1 (IA generativa y retroalimentación formativa) y los negros a C2 (reflexión pedagógica). PE = prompt estructurado. FF = feedforward recibido. DI = diálogo iterativo. VC = verificación crítica. AD = alucinación detectada. AC = andamiaje cognitivo. RD, RA y RC = reflexión descriptiva, analítica y crítica. MC = metacognición. RE = rediseño didáctico. GT = deliberación en el GET. AU = autonomía profesional.

DISCUSIÓN

La asociación del prompt estructurado con el feedforward y de este con el rediseño reproduce la lógica de una retroalimentación orientada al paso siguiente (Hattie & Timperley, 2007). En dichos términos, Hofmann et al. (2025) exploraron la retroalimentación generada por IA como una vía hacia las habilidades reflexivas, pero la revisión de Zhai et al. (2024) previene contra una dependencia que erosionaría las capacidades cognitivas. Los resultados ubican ambas posturas en un mismo escenario, ya que la verificación crítica tuvo una ganancia junto con la reflexión analítica y coocurrió con la detección de alucinaciones. La diatriba se resolvió mediante un procedimiento que obligó a contrastar y que trasladó la decisión final al practicante.

Steiss et al. (2024) y Banihashem et al. (2024) coinciden en describir los perfiles distintos para la retroalimentación humana y la de IA. De ahí que el diseño no sustituye al formador. La deliberación en el GET vinculada aquí con la reflexión crítica aportó la necesaria sensibilidad socioemocional que la IAG no ofrece. Este es un rasgo que ninguno de los estudios señalados aborda por su foco en la escritura. Una vertiente crítica recuerda que la adopción de la IA en la educación arrastra supuestos que conviene examinar en detalle posteriormente (Williamson & Eynon, 2020), y la síntesis de Celik et al. (2022) equilibró las promesas y desafíos para los docentes. El desarrollo acá presentado responde a estos aportes al tratar la salida de la IAG como el objeto de estudio y no como un veredicto.

Según las fuentes revisadas, las investigaciones sobre los procesos reflexivos en la práctica (García Lazaro & Reyes de Cózar, 2024) y sobre los usos docentes de la IAG (Martínez-Garcés et al., 2026) avanzan por rutas separadas, y los estudios en formación inicial exploran la escritura (Ossa & Willatt, 2023) o la planificación (Farillon-Labis, 2025) sin llegar a integrar el prompting estructurado, la deliberación grupal, los ciclos de investigación-acción y el seguimiento del tránsito entre los niveles de reflexión. Los hallazgos reflejaron un alcance de esa integración en cuatro semanas, por lo que se constituye en un formato replicable en escuelas pedagógicas de práctica intensiva; y se inscribe en la agenda general de investigación sobre IA en educación esbozada por Hwang et al. (2020).

Dos estudios con docentes en formación (Hofmann et al., 2025; Farillon-Labis, 2025) sirven para encontrar un contrapunto. Hofmann et al. (2025) se ocuparon de las habilidades reflexivas y Farillon-Labis (2025) de las reflexiones durante la planificación de clases. En ese par de obras la IAG se posicionó antes de la sesión. El modelo acá aplicado conservó ese momento previo y añadió además la reflexión posterior a la ejecución en el aula, con deliberación grupal, para cerrar el bucle entre la sugerencia y la práctica. Liu et al. (2025) estudiaron la intención de adopción como una variable previa a la conducta, mientras que lo acá desarrollado observó las conductas de uso como la estructura de las peticiones y la verificación de las salidas. La indagación colaborativa para la alfabetización en IA (Kelley & Wenzel, 2025) comparte con este trabajo la lógica de aprender en comunidad pero aclarando que aquí el foco recayó sobre la reflexión pedagógica.

La exigencia de declarar la traza de interacción con la IAG se articula con el marco de política institucional de Chan (2023). Otros estudios como el de Ceballos Macías (2025) y el aporte de Sandoval Hernández et al. (2024) reportaron usos de la IAG en la docencia, y López Tineo et al. (2025) describieron asistentes virtuales, de modo que la ruta acá aplicada incorporó el análisis del proceso reflexivo. La disposición de los futuros docentes a adoptar la IAG (Liu et al., 2025) y la alfabetización por indagación colaborativa (Kelley & Wenzel, 2025) ofrecen condiciones favorables para replicar el ciclo.

Se observó a lo largo del proceso que el papel del formador cambió de corrector a un interlocutor del proceso. Su tiempo se liberó de la tarea de la revisión descriptiva y se

reorientó hacia la conversación sobre las decisiones didácticas, lo cual coincidió con lo declarado en las entrevistas. En términos de la protección de datos, el protocolo prohibió ingresar información identificable de los estudiantes y exigió declarar la traza de interacción, medidas estas que respondieron al riesgo de exposición de los datos y de plagio algorítmico. La prohibición se sustentó en la disciplina individual, de tal forma que la institución debería ofrecer plataformas con garantías de privacidad y lineamientos claros, en la línea del marco de política institucional de Chan (2023).

La revisión de Martínez-Garcés et al. (2026) recogió también las cuestiones éticas asociadas con el uso de la IAG en la docencia. El trabajo que se desarrolló tradujo esa preocupación en unos procedimientos concretos como fueron el consentimiento informado, la anonimización de los datos, la verificación de las salidas y la declaración de la traza de interacción. Esa traducción resultó útil para la enseñanza de la investigación, ya que permitió que los practicantes viesen cómo un principio ético se convirtió en un instrumento de registro y en un criterio de evaluación.

Para la réplica en una fase piloto se recomienda mantener dos ciclos breves, fijar de antemano la rúbrica de reflexión, y registrar la versión del modelo de IAG usada por cada practicante, así como el reservar un tiempo de deliberación en cada GET. Esas decisiones ayudan a proteger la comparabilidad de los registros y reducen el riesgo de que la variabilidad tecnológica oculte los cambios reflexivos.

Entre los aportes de la investigación-acción destacaron la participación de los practicantes en el diseño de las mejoras, la retroalimentación inmediata antes de la sesión, el contraste explícito entre las versiones del plan y la construcción colectiva de criterios en el GET. Las limitaciones pedagógicas son de peso. La IAG carece de sensibilidad hacia la dinámica sociocultural del aula y puede ofrecer sugerencias estandarizadas. Cuatro semanas impiden observar la consolidación de la autonomía. El muestreo intencional limitó la transferibilidad y la variabilidad entre las versiones gratuitas y de pago de los modelos, generando una experiencias desiguales. Cabe añadir que la rúbrica de reflexión, aunque fue aplicada con doble codificación, redujo la complejidad del pensamiento docente a tres niveles, y que la comparación entre el plan inicial y el reestructurado no aisló el efecto de la IAG del efecto de la deliberación grupal ni del paso del tiempo.

CONCLUSIONES

Como respuesta al objetivo general, el modelo indicó que la retroalimentación con IAG, estructurada con los marcos PARTS y CLEAR y con la Escalera de Wilson e integrada en dos ciclos breves de investigación-acción, favorece el tránsito de la reflexión descriptiva hacia la analítica y la crítica, y puede acompañar positivamente la reestructuración didáctica de los planes de sesión. Esa posibilidad depende de la verificación crítica de las salidas y de la deliberación en el GET. En esos términos, la IAG ocupó el rol de un interlocutor sometido al juicio del grupo y no el de una autoridad evaluadora.

En el plano institucional, desde el modelo aplicado se sugiere que las escuelas pedagógicas den carácter institucional a los GET y al registro de la traza de interacción con la IAG, y que reserven al formador la retroalimentación socioemocional y la decisión final. La revisión ética reveló la necesidad de exigir el consentimiento, la anonimización, y también la prohibición de ingresar datos de escolares y la declaración del uso de la IAG. Desde el plano curricular, se invita a incorporar la alfabetización crítica en IA, con talleres de prompting, dentro de la práctica preprofesional; acompañando así lo pedagógico, donde se propone que el practicante discuta, rechace o reformule cada sugerencia y documente sus decisiones.

CC BY NC SA

REFERENCIAS

Banihashem, S. K., Taghizadeh Kerman, N., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*, Artículo 23. https://doi.org/10.1186/s41239-024-00455-4

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328-352. https://doi.org/10.1080/14780887.2020.1769238

Ceballos Macías, J. D. (2025). Impacto de la inteligencia artificial en la formación docente en el CUNORTE. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 16*(31), e998. https://doi.org/10.23913/ride.v16i31.2695

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*, 616-630. https://doi.org/10.1007/s11528-022-00715-y

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00408-3

Farillon-Labis, S. J. (2025). Exploring pre-service mathematics teachers' reflections in lesson planning with generative artificial intelligence. *International Journal of Learning, Teaching and Educational Research, 24*(11), 334-355. https://doi.org/10.26803/ijlter.24.11.16

García Lazaro, I., & Reyes de Cózar, S. (2024). Preservice teachers' reflective processes during field experiences: A systematic review from the ONION model framework. *Reflective Practice*. https://doi.org/10.1080/14623943.2024.2421597

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hofmann, F., Daunicht, T.-M., Plößl, L., & Gläser-Zikuda, M. (2025). Promoting reflection skills of pre-service teachers: The power of AI-generated feedback. *Education Sciences, 15*(10), Artículo 1315. https://doi.org/10.3390/educsci15101315

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Artículo 100001. https://doi.org/10.1016/j.caeai.2020.100001

Kelley, M., & Wenzel, T. (2025). Advancing artificial intelligence literacy in teacher education through professional partnership inquiry. *Education Sciences, 15*(6), Artículo 659. https://doi.org/10.3390/educsci15060659

Liu, Y., Wang, Q., & Lei, J. (2025). Adopting generative AI in future classrooms: A study of preservice teachers' intentions and influencing factors. *Behavioral Sciences, 15*(8), Artículo 1040. https://doi.org/10.3390/bs15081040

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*(4), Artículo 102720. https://doi.org/10.1016/j.acalib.2023.102720

López Tineo, N. K. V., Reyes Tejada, G. S., Wong Lam, L. J., & Llanco Gonzales, D. I. (2025). Implementación de asistentes virtuales basados en inteligencia artificial generativa creados en el Laboratorio EDU-INNOVA para mejorar la enseñanza de los docentes. *Desde el Sur, 17*(2), Artículo e0020. https://doi.org/10.21142/des-1702-2025-0020

Martínez-Garcés, J., Asprilla-Castillo, V., & Montenegro-Perafán, E. (2026). Inteligencia artificial generativa en la pedagogía universitaria. Revisión sistemática. *Alteridad, 21*(2), 207-221. https://alteridad.ups.edu.ec/alteridad/article/view/12979

Ossa, C., & Willatt, C. (2023). Providing academic writing feedback assisted by generative artificial intelligence in initial teacher education contexts. *European Journal of Education and Psychology, 16*(2), 1-16. https://doi.org/10.32457/ejep.v16i2.2412

Sandoval Hernández, M. A., Morales Alarcón, G. J., Vázquez Leal, H., Huerta Chua, J., & Filobello Niño, U. A. (2024). El uso del prompt de ChatGPT como asistente en la educación. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 14*(28), e645. https://doi.org/10.23913/ride.v14i28.1872

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, Artículo 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223-235. https://doi.org/10.1080/17439884.2020.1798995

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*. https://doi.org/10.1186/s40561-024-00316-7