

López Gil, K., & Moreno Mosquera, E. (2025, enero-abril). Retroalimentación formativa en la escritura de tesis en posgrado: Comparación entre ChatGPT y revisores pares en un círculo de escritura. *Revista Virtual Universidad Católica del Norte*, (74), 123-160.

<https://www.doi.org/10.35575/rvucn.n74a6>

**Retroalimentación formativa en la escritura de tesis en posgrado:
Comparación entre ChatGPT y revisores pares en un círculo de escritura**

Formative Feedback in Graduate Thesis Writing: A Comparison between ChatGPT and Peer Reviewers in a Writing Group

Karen López Gil

Doctora en Educación
Facultad de Humanidades, Universidad del Valle
Cali, Colombia

karen.lopez.gil@correounivalle.edu.co

Orcid: <https://orcid.org/0000-0001-9826-0799>

CvLAC:

https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0001402978

Emilce Moreno Mosquera

Doctora en Educación
Facultad de Educación, Pontificia Universidad Javeriana
Bogotá, Colombia

moreno-e@javeriana.edu.co

Orcid: <https://orcid.org/0000-0003-4058-5691>

CvLAC:

https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000825352

Recibido: 12 de abril de 2024

Evaluado: 21 de agosto de 2024

Aprobado: 21 de noviembre de 2024

Tipo de artículo: Investigación.

Resumen

La evaluación formativa en escritura académica es reconocida por su potencial tanto para la mejora de los textos como para el desarrollo de los escritores. Este artículo evalúa los efectos de la retroalimentación proporcionada por ChatGPT y por revisores pares a la escritura de



introducciones de tesis en un círculo de escritura en posgrado. Se llevó a cabo una investigación de enfoque mixto, con un diseño concurrente de alcance exploratorio. La recolección de información incluyó 12 borradores de texto, los comentarios de los pares y de ChatGPT a partir de rúbricas, las versiones ajustadas de las introducciones y las percepciones de los participantes sobre la calidad, pertinencia e impacto de la retroalimentación. Los resultados mostraron congruencia entre las calificaciones proporcionadas por ambas fuentes y aceptación de la mayoría de las sugerencias. Se identificó que ChatGPT se centró en la estructura del texto y en aspectos estilísticos, mientras que los pares se enfocaron en los contenidos y en la viabilidad de las propuestas de investigación. Los hallazgos evidencian el potencial educativo de ChatGPT en la personalización del aprendizaje de la escritura al proporcionar apoyos eficientes y oportunos, y lo sitúan como una estrategia complementaria a la retroalimentación humana.

Palabras clave: ChatGPT; Escritura; Inteligencia artificial; Posgrado; Retroalimentación.

Abstract

Formative assessment in Academic Writing is recognized for its potential to improve both texts and writers' development. This research evaluated the effects of feedback provided by ChatGPT on the writing of thesis introductions within the context of a Postgraduate Writing Group. A mixed-methods study was conducted, with a concurrent design of exploratory scope. Data collection included 12 text drafts, peer and ChatGPT feedback based on rubrics, revised versions of the introductions, and participants' perceptions regarding the relevance, quality and impact of the feedback. Findings indicated consistency in the ratings provided by both reviewers, with over half of the suggestions being incorporated into the introductions. Despite the use of a rubric to share evaluation criteria, it was observed that ChatGPT focused on text structure and formal aspects of writing, whereas peers concentrated on content development and the feasibility of research proposals. The findings show the educational potential of ChatGPT in personalizing writing learning by providing efficient and timely support and position it as a complementary strategy to human feedback.

Key words: ChatGPT; Writing; Artificial intelligence; Postgraduate; Feedback.

Introducción

La retroalimentación formativa ha sido ampliamente reconocida por su enfoque efectivo para mejorar el aprendizaje de los estudiantes, en tanto aporta a la disminución de las brechas entre el desempeño actual y la meta deseada (Narciss & Zumbach, 2020). Una retroalimentación de calidad,

No solo tiene que informar de los errores, sino también proveer información suficiente y oportuna (...) para favorecer el cambio a lo largo de un proceso de aprendizaje, destacando los logros del aprendiz y lo que debe repensar o rehacer. (Cuevas Solar & Arancibia, 2020, p. 35)

Sin embargo, debido a la limitación de los recursos, en la mayoría de los sistemas educativos es difícil que los docentes puedan proporcionar retroalimentación oportuna y continua a todos sus estudiantes. Esta situación es todavía más desafiante en procesos complejos y de largo alcance como la escritura de tesis en los niveles de grado y posgrado. Por ello, en las instituciones de educación superior se han planteado distintas alternativas de apoyo a los estudiantes, como es el caso de las tutorías en centros de escritura (Molina & López-Gil, 2020), las revisiones entre compañeros en los grupos de escritura (Colombo et al., 2022) o, más recientemente, opciones de retroalimentación con herramientas apoyadas en inteligencia artificial (IA).

El uso de las tecnologías para la evaluación de la escritura es de larga data. La evaluación automatizada de la escritura (EAES) (Strobl et al., 2019) surgió como campo interdisciplinario en la década de los 60, inicialmente centrado en la calificación de pruebas de escritura en inglés como lengua extranjera. Sus propósitos se han ido extendiendo y actualmente se emplea tanto con fines sumativos como para apoyar el aprendizaje de los estudiantes (Lillo-Fuentes et al., 2023).

A partir del desarrollo de la IA, la mayoría de las herramientas de evaluación automatizada se han ido transformando. Ding y Zou (2024) afirman que herramientas de escritura apoyadas en IA, como Grammarly, QuillBot o Wordtune, pueden incidir en el desarrollo de habilidades de escritura de los estudiantes. También pueden aportar al contenido y a la estructuración de los textos (Castillo-González, 2023; Hill-Yardin et al., 2023; Zhai & Ma, 2023). En el funcionamiento de estas tecnologías, destaca el uso de algoritmos avanzados para identificar errores comunes en gramática, puntuación, sintaxis y estilo (Tambunan et al., 2022) y opciones como parafrasear y

refinar oraciones, que pueden convertir el proceso de escritura en una oportunidad de aprendizaje y de mejora continua.

Sin embargo, en el proceso de retroalimentación estas herramientas pueden no ser tan efectivas para abordar elementos de escritura de orden superior, como la estructura y la coherencia de los argumentos. Estos aspectos requieren de un análisis profundo del tema, pensamiento lógico y la conexión de ideas, que actualmente están más allá de la capacidad de las herramientas (Farrokhnia et al., 2023). De igual modo, se identifica como limitación que la mayoría de las tecnologías se han propuesto para la escritura en el idioma inglés, si bien hay algunas experiencias de uso en español como arText, un sistema automático de ayuda en la redacción (Núñez & da Cunha, 2022) o Estilector, un sistema automático de retroalimentación escrita (Nazar & Renau, 2023).

Con los recientes avances en los grandes modelos de lenguaje (LLM por sus siglas en inglés), en particular con la masificación al acceso de aplicaciones interactivas de chatbot como ChatGPT, se están generando altas expectativas en relación con sus usos en el ámbito educativo, sobre todo por su capacidad de analizar los textos de manera más global, de producir textos originales, de conservar ciertos datos en la memoria y de hacer ajustes a partir de la interacción con el usuario. Esto ayudaría a superar las limitaciones tradicionales de las herramientas de EAES que suelen enfocarse en aspectos superficiales del texto y que suelen requerir extensos periodos de entrenamiento, además de altos costos (Naismith et al., 2023). En esta línea, se plantea que ChatGPT puede contribuir con el aprendizaje personalizado, en la medida en que puede adaptar los contenidos y la instrucción para responder a las necesidades y preferencias de los estudiantes, lo que ayudaría no solo con la mejora en las dimensiones de los productos escritos, sino del mismo proceso de escritura (Atlas, 2023).

Si bien estas herramientas tienen capacidades emergentes que pueden aprovecharse en procesos escriturales, no han sido creadas específicamente para la evaluación automatizada de la escritura. Programas como Grammarly, por ejemplo, están diseñados para puntuar y retroalimentar textos basándose en grandes corpus modelo para determinar si la escritura de un estudiante es estadísticamente similar a la de esos modelos. Los LLM no están alimentados con corpus de un dominio específico, sino con información extraída de internet. La naturaleza de dominio general

representa ventajas y desventajas en el proceso de retroalimentación, por lo que es necesaria su exploración de manera sistemática.

Esta investigación busca una mejor comprensión sobre cómo se puede usar la IA generativa, específicamente ChatGPT, como herramienta de retroalimentación formativa de la escritura de trabajos de investigación. La escritura de la tesis se ha identificado como uno de los factores académicos más desafiantes para la culminación exitosa de los estudios posgraduales (Gómez & Alzate, 2018); incluso, se ha acuñado un término para describir este fenómeno: “Todo menos la tesis” (en inglés, *All But Dissertation*) (Abreu, 2015; Godoy & Damián Ferro, 2020). De ahí que las instituciones de educación superior estén avanzando en el desarrollo de estrategias de acompañamiento a la escritura, con el fin de disminuir las tasas de deserción y de rezago educativo (Castro, 2024). En este panorama, esta investigación se plantea como objetivo evaluar los efectos de la retroalimentación proporcionada por ChatGPT y por revisores pares a la escritura del apartado introductorio de tesis de posgrado. Para ello, se proponen como preguntas específicas de la investigación: ¿hay congruencias en las puntuaciones asignadas por ChatGPT y por los revisores pares?, ¿qué tipos de sugerencias predominan en cada una de las fuentes evaluativas?, ¿cuál es el nivel de aceptación, por parte de los escritores, de las sugerencias planteadas por ChatGPT y los revisores pares?, ¿cuáles son las percepciones de los escritores en términos de la pertinencia, calidad e impacto de la retroalimentación de ambas fuentes en sus procesos de escritura?

La escritura se constituye en uno de los principales vehículos para la construcción del conocimiento y para la validación de los aprendizajes en la educación superior. Es crucial para desarrollar y evaluar el pensamiento crítico, la argumentación, la comprensión de la información, el dominio de la lengua, entre otros aspectos. La integración de las tecnologías en los procesos de escritura es casi inevitable, pero no puede darse de forma acrítica. Por esta razón, es fundamental desarrollar investigaciones que analicen su impacto y sus posibles implicaciones en el contexto educativo. Esto permitirá que los educadores tomen decisiones reflexivas e informadas sobre cómo pueden emplear (o no) estas herramientas en los procesos pedagógicos.

ChatGPT como evaluador formativo de la escritura

Las IA generativas de textos son sistemas avanzados de procesamiento natural del lenguaje que están diseñados para crear contenidos escritos de manera autónoma. Estas IA se basan en grandes modelos de procesamiento de lenguaje natural (LLM), entrenados con enormes cantidades de datos textuales que proporcionan la infraestructura y los algoritmos fundamentales para que las IA puedan comprender y generar escritos coherentes y contextualmente relevantes (Meyer et al., 2023). Actualmente, LLM como GPT-3.5, GTP-4 y GPT-4.o de OpenAI, Gemini Pro de Google, LLaMa de Meta y Claude v3 de Anthropic fundamentan la mayoría de las aplicaciones de generación de textos.

ChatGPT es una aplicación de chatbot basada en los LLM GPT, que se caracteriza por interactuar conversacionalmente con el usuario y aprender por refuerzo a partir de la interacción con los humanos (OpenAI, 2023). Al apoyarse en grandes cantidades de información y producir textos coherentes, se ha considerado de gran potencial en la escritura.

Desde que inició el acceso masivo a la aplicación de ChatGPT a finales del 2022, la incursión en el ámbito educativo no se ha hecho esperar y se han desarrollado múltiples estudios alrededor de este fenómeno. Los primeros temores de docentes e investigadores se relacionaban con su capacidad de producción de textos originales y cómo esto podría terminar por reemplazar la escritura de los estudiantes. Por esta razón, se han estudiado las semejanzas y diferencias entre las producciones de alumnos universitarios y ChatGPT (Herbold et al., 2023), se ha pedido a la herramienta responder a exámenes especializados (Dergaa et al., 2023), apoyar procesos de escritura creativa (Bierman et al., 2022; Lozić & Štular, 2023), entre otras tareas. Las investigaciones han identificado también las limitaciones de ChatGPT y los desafíos que plantea a la educación, en particular respecto a la producción de géneros académicos tradicionales como los resúmenes, las reseñas y los ensayos (Imran & Almusharraf, 2023). Entre los principales riesgos se encuentran los problemas de integridad académica, plagio, información imprecisa y desinformación, razón por la cual se ha considerado necesario el rediseño de la evaluación escrita y el establecimiento de políticas de uso tanto en el ámbito académico como en el científico (Farazouli et al., 2023; Hsiao et al., 2023).

Una de las aplicaciones de ChatGPT que se ha explorado en el contexto educativo es su posible rol en procesos de evaluación formativa. Los chatbots se han identificado como herramientas efectivas de retroalimentación en el componente escritural en el aprendizaje de idiomas (Algaraady & Mahyoob, 2023; Jiang et al., 2023; Yeşilyurt & Sezgin, 2023), al mostrar efectividad en aspectos como la identificación de errores de concordancia sujeto-verbo, formas verbales, omisión de determinantes, repetición de palabras, redundancias, etc. Mizumoto y Eguchi (2023) evaluaron un corpus de 12.000 ensayos de escritores no nativos en inglés, con una herramienta de puntuación automatizada apoyada en GPT-3, y encontraron viabilidad y confiabilidad en el uso de estas tecnologías.

Asimismo, se ha identificado consistencia entre las calificaciones de la IA y las asignadas por evaluadores humanos. Hackl et al. (2023) destacan que esta consistencia es mayor en criterios relacionados con el estilo y la precisión de los contenidos en escritos cortos. Naismith et al. (2023) encontraron que ChatGPT puede producir calificaciones comparables con las humanas en aspectos como la coherencia del discurso. Guo & Wang (2024), por su parte, evidenciaron que las evaluaciones suelen tener distintos focos: los profesores tienden a centrar su evaluación en aspectos del contenido, mientras que ChatGPT atiende de manera similar contenido, estructura y uso del lenguaje; adicionalmente, ChatGPT suele proporcionar una retroalimentación más directa que la humana, con recomendaciones específicas sobre cómo ajustar el texto.

Escalante et al. (2023) hicieron mediciones sucesivas del efecto de la retroalimentación humana y de IA en la escritura académica y hallaron impactos similares en la reescritura. También, exploraron las percepciones de los participantes y encontraron una preferencia ligeramente a favor de la retroalimentación humana. Cabe destacar que en esta experiencia los participantes no interactuaron directamente con ChatGPT, sino que recibieron sus observaciones a través del profesor. Dai et al. (2023) usaron ChatGPT para proporcionar retroalimentación en escritos de estudiantes universitarios. En este caso, se encontró que los comentarios de la IA eran más legibles y detallados que los comentarios de los instructores y que había cierto grado de acuerdo en las evaluaciones, pero no encontraron consistencia en todos los criterios. En investigaciones más recientes, Steiss et al. (2024) compararon la retroalimentación proporcionada por humanos expertos y por ChatGPT a ensayos escritos de estudiantes de secundaria. Si bien encontraron ligeras diferencias en términos de la calidad de la retroalimentación, a favor de la revisión humana,

concluyen que estas diferencias son modestas al considerar la calidad general y el ahorro de tiempo de las herramientas apoyadas en IA. Banihashem et al. (2024) contrastaron las retroalimentaciones a ensayos argumentativos en posgrados e identificaron una mayor calidad en las revisiones de los pares que de la IA generativa. No obstante, concluyen que las evaluaciones automatizadas son pertinentes y pueden emplearse de manera complementaria.

Finalmente, un antecedente interesante para esta investigación lo constituyó el estudio de Tzirides et al. (2023), en el que se comparó la retroalimentación proporcionada por IA (modelo GPT-3.5) y la retroalimentación por pares en la producción de textos multimodales. Aunque identificaron algunos aportes, los autores son menos entusiastas respecto al potencial de los LLM en la retroalimentación de la escritura especializada. De igual modo, encontraron valoraciones menos positivas de los participantes, quienes identificaban en la herramienta problemas de atribución del sentido global de los textos, sugerencias muy generales o de naturaleza únicamente estilística.

Estas investigaciones, bastante recientes y algunas todavía en construcción, muestran un campo emergente en la retroalimentación automática de la escritura. Los resultados evidencian varias ventajas de la IA en esta tarea, en las que se destacan la inmediatez de respuesta, la posibilidad de ampliación de la información y la interacción conversacional con el usuario. Destacan la necesidad de “entrenar” las herramientas con modelos de texto y de proporcionar criterios detallados para la evaluación a través de rúbricas y otros instrumentos. Asimismo, coinciden en la relevancia de los *prompts* o indicaciones para guiar a ChatGPT hacia los comportamientos deseados especificando propósitos, contexto, resultados, estilo y formato de las tareas. Aunque las investigaciones coinciden en varios aspectos, también se identifican discrepancias en relación con la consistencia de las evaluaciones entre humanos y máquinas, la valoración de la calidad de la retroalimentación y la favorabilidad o aceptación de la retroalimentación por parte de los escritores. Este último es un aspecto esencial, en tanto el éxito de cualquier tecnología de aprendizaje depende de si los usuarios la adoptan. Una limitación identificada en estos estudios es que se centran en escrituras escolarizantes y generales como la producción de ensayos argumentativos en secundaria, en primeros años de universidad o en una lengua extranjera, dejando de lado el trabajo con textos especializados, como es el caso de las tesis de posgrado en disciplinas particulares.

Integración de IA en los círculos de escritura

Los círculos o grupos de escritura hacen parte de los dispositivos de acompañamiento a la escritura en nivel posgradual. Son comunidades colaborativas de escritores que se reúnen para desarrollar y revisar sus producciones (Colombo et al., 2022). Se conciben como lugares inspiradores y creativos donde la gente habla, escribe y aprende junta porque están siendo nutridos, empoderados y estimulados (Aitchison, 2009). Son espacios 'horizontalizados', debido a su composición orientada a los estudiantes, quienes son acompañados por un facilitador experto en escritura. A través de la retroalimentación se permite que los estudiantes sean más conscientes de los apartados y aspectos particulares de la tesis, las normas y las convenciones de la escritura académica en investigación (Aitchison, 2009).

Escribir a nivel de maestría y doctorado representa para los estudiantes el ingreso a su comunidad disciplinar; en ese sentido, la escritura se va configurando como un dispositivo fundamental para que desarrollen su identidad académica y profesional (Cuthbert et al., 2009; Lillis & Scott, 2015). Formar parte de una comunidad científica supone conocer las convenciones de las disciplinas, de ahí la necesidad de involucrar a los estudiantes en prácticas de enseñanza de géneros especializados, prácticas que toman sentido en el uso del código escrito en el marco de una cultura, lo cual implica considerar las ideologías, las identidades y las estructuras sociales y disciplinares del escritor. Esto es, la literacidad comprendida en el marco de la metáfora de la ecología (Barton & Hamilton, 2004). En los círculos de escritura, los estudiantes están inmersos en prácticas de revisión y ajuste que no solo favorecen su participación en la comunidad disciplinar, sino que también propician una gestión más eficiente del proceso de escritura. Así mismo, el afianzamiento de las prácticas de escritura se ve beneficiado por el compromiso social, dado que los grupos se constituyen en una forma poderosa de escribir en un trabajo colaborativo, que a su vez da herramientas para mejorar la propia capacidad de producir textos.

En el contexto de los círculos de escritura, el papel del docente como única fuente de evaluación se ha desplazado y se reconocen las ventajas de la creación de espacios colaborativos, de discusión y de desarrollo de la autonomía y la autorregulación (Colombo et al., 2022; Espeche et al., 2024). En ese sentido, los círculos son entornos propicios para preguntarse por el posible rol

de las tecnologías, en particular de la IA generativa, en procesos de retroalimentación de la escritura, en los que podría funcionar como otro “agente” o “socio” en las revisiones o, como con las tecnologías antecesoras, como una herramienta de apoyo. Se parte de apuestas teóricas socioculturales y críticas sobre la mediación de las tecnologías digitales en la escritura académica, que reconocen, más allá de los avances y características de las herramientas, el papel de los usos, representaciones, valoraciones y sentidos atribuidos por los sujetos en el marco de prácticas sociales situadas (Barton & Lee, 2013; Street, 2017).

Método

Enfoque de investigación

Esta investigación fue de enfoque mixto, en tanto articuló propósitos, técnicas de recolección de datos y técnicas de análisis cuantitativas y cualitativas. Se priorizó el enfoque cualitativo al estudiar las prácticas de revisión y retroalimentación de un grupo reducido de estudiantes, pero se emplearon datos cuantitativos para reconocer los usos concretos y cambios en los textos. De esta manera, se planteó un diseño concurrente en el que se recogieron y analizaron de forma paralela los datos (Creswell & Plano, 2018). Al ser el estudio de la IA generativa y su vínculo con la escritura académica un campo todavía incipiente, se definió un alcance exploratorio para la investigación.

Muestra

La investigación se desarrolló en el marco de un círculo de escritura conformado por 12 estudiantes de tercer semestre de una maestría en educación. Se hizo una convocatoria abierta para que se inscribieran voluntariamente personas interesadas en avanzar en la escritura del apartado introductorio de la tesis a través de procesos de retroalimentación. El círculo se constituyó como un espacio alternativo y complementario a la asesoría de tesis y contó con la coordinación logística de dos docentes, una especialista en escritura académica y otra en investigación educativa. En este

caso, en la investigación se optó por un muestreo no probabilístico por conveniencia, que incluyera la totalidad de los participantes del círculo de escritura.

Selección de la IA generativa

Para elegir la IA generativa, se compararon tres herramientas basadas en los LLM con mayor desarrollo en el periodo 2023-2: ChatGPT en sus versiones gratuita (basado en GPT 3.5) y plus (GPT 4), Bard de Google (basado en Gemini Pro) y Claude Plus de Anthropic (versión 2). Las tres aplicaciones cumplieron con las características requeridas para procesos de retroalimentación escrita: disponibilidad de interactuar en español, instrucciones a través de *prompts* escritos, lectura de imágenes y cuadros, procesamiento de documentos en distintos formatos (versiones plus), acceso a datos actualizados a través de internet (versiones plus) y soporte de al menos 3000 palabras por entrada.

Se eligió ChatGPT Plus, pues, además de cumplir con las características requeridas, se basa en el LLM GPT-4, el modelo de procesamiento que fundamenta la mayoría de las aplicaciones de IA generativas textuales, lo que permite a los estudiantes una mayor transferencia de los aprendizajes sobre la aplicación. De igual modo, los participantes estaban ya familiarizados con la herramienta y la institución educativa contaba con suscripciones a la versión de pago.

Proceso de investigación

La investigación se desarrolló en tres fases. La primera correspondió al diseño de la propuesta de retroalimentación. Para ello, se organizaron las indicaciones y se creó una rúbrica con los criterios de evaluación para emplear en las revisiones, que se introdujeron como *prompts* o instrucciones en ChatGPT. Para facilitar el proceso, las coordinadoras modelaron la retroalimentación con tres introducciones de tesis de maestría que se usaron como ejemplos. Se solicitó a cada estudiante que practicara con estos ejemplos en ChatGPT para entrenar la IA y para mantener un registro de la revisión.

Cabe mencionar que el propósito del círculo de escritura no era hacer una corrección de estilo o edición de los documentos, sino ayudar a los participantes a identificar las características

de sus textos y a reflexionar sobre sus procesos escriturales. Por ello, las indicaciones se centraron en hacer comentarios y sugerencias y no en modificar directamente los escritos. Si bien los *prompts* podían tener ajustes de acuerdo con las necesidades de los participantes y del momento de la revisión, la instrucción inicial para ChatGPT fue:

"Teniendo en cuenta los tres ejemplos que acabamos de revisar, te proporcionaré la introducción de una tesis de maestría en educación sobre [breve mención del tema] y una rúbrica de evaluación con diez criterios y cinco niveles de desempeño. Tu tarea es revisar el texto en función de la rúbrica y ofrecer retroalimentación detallada y constructiva. Por favor, evalúa cada criterio de la rúbrica de manera individual, justificando el nivel de desempeño obtenido. Proporciona comentarios puntuales, asegurándote de señalar los puntos fuertes y áreas que requieren ajustes, con sugerencias específicas para mejorar cada parte de la introducción.

Aquí está la rúbrica: [Introducir el texto aquí]

Aquí está la introducción: [Introducir el texto aquí]"

Se llevó a cabo una sesión de acercamiento de los participantes a las dinámicas de revisión entre pares para que pudieran comprender el sentido y estrategias efectivas de retroalimentación formativa. De igual manera, se reservó una sesión para explorar las características de ChatGPT en su versión Plus (GPT-4), conocer su funcionamiento y discutir sobre sus alcances y limitaciones. La formación fue impartida por una experta en escritura académica y en procesos de retroalimentación escrita, con experiencia en el empleo de herramientas automatizadas de evaluación de la escritura.

En la segunda fase, implementación, se llevaron a cabo las revisiones. Se destinaron cuatro encuentros para la revisión de la introducción y del marco teórico. Sin embargo, en este artículo se presentan los hallazgos de la primera revisión, debido a que en ella se generó la mayor cantidad de comentarios, mientras que en las siguientes se hizo seguimiento de los ajustes. Cada participante presentó un borrador de su introducción de acuerdo con la guía de trabajo de grado del programa de maestría, que establecía un rango aproximado de 2000-2500 palabras más referencias. En las revisiones entre pares se diligenciaron las rúbricas y se hicieron comentarios específicos en el texto; mientras que las revisiones con ChatGPT incluyeron la rúbrica y preguntas adicionales que los escritores hicieron a la herramienta. Posterior a la revisión, los participantes dieron respuesta a

cada comentario e incorporaron los ajustes que consideraron relevantes en una nueva versión del texto.

En la última fase, evaluación, se analizó la consistencia entre las retroalimentaciones de las dos fuentes, se identificaron los cambios incorporados en las nuevas versiones de los textos y se exploraron, a través de grupos focales, las percepciones de los participantes respecto a la retroalimentación y su impacto tanto en los textos como en sus procesos de escritura.

Recolección de datos

El registro de la información se hizo a través de varios mecanismos: rúbricas de evaluación, revisión documental y grupos focales.

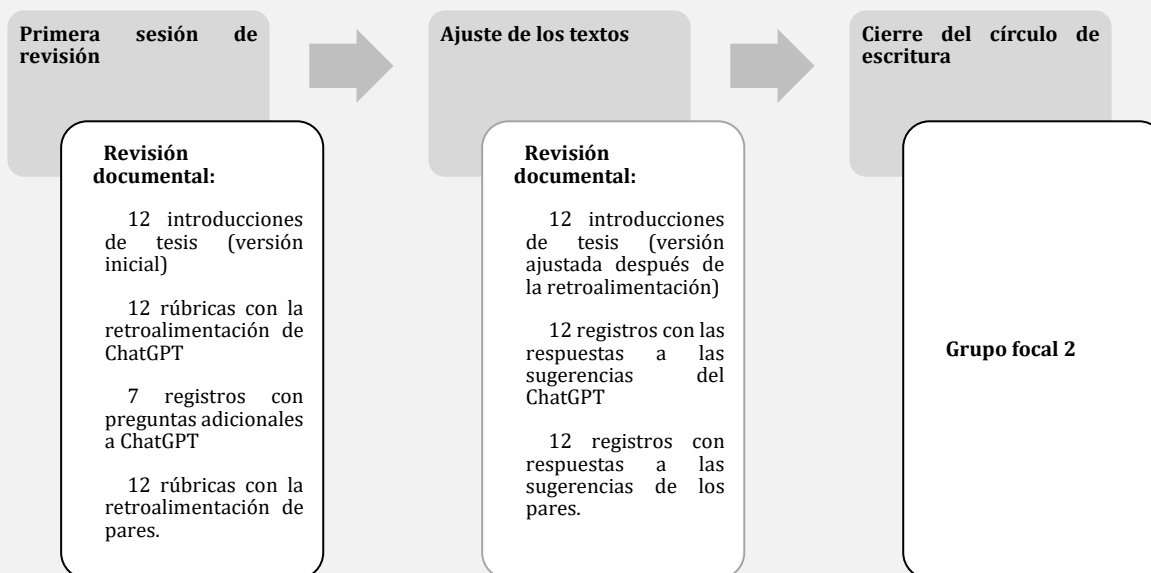
Para identificar la congruencia entre las valoraciones proporcionadas por ChatGPT y por los revisores pares, se recogieron las rúbricas diligenciadas por ambas fuentes de retroalimentación. Estas rúbricas estaban organizadas a partir de 10 criterios establecidos por el programa de Maestría para evaluar el apartado introductorio de la tesis: contextualización del objeto de estudio, delimitación del problema de investigación, revisión de literatura, planteamiento de los objetivos, justificación, coherencia, cohesión, estilo académico formal, corrección gramatical y ortográfica, y formato de citación con estilo APA 7^a. edición. Cada criterio incluyó cinco niveles de desempeño con sus descriptores, con una escala de calificación numérica en un rango de 0 a 5, siendo 5 la puntuación máxima. Además de la puntuación, la rúbrica incluyó una columna en la que se debía justificar la nota asignada, así como las recomendaciones cuando se requirieran ajustes. En siete casos, los autores incluyeron mecanismos adicionales de retroalimentación de ChatGPT, como transformar la rúbrica en listas de chequeo con preguntas específicas, o solicitudes de corrección de estilo.

Para la revisión documental, se recogieron las 12 versiones iniciales y las 12 versiones ajustadas de las introducciones, así como las respuestas y reacciones a los comentarios de los revisores. Con esta técnica se buscó identificar el impacto de la retroalimentación proporcionada por ambas fuentes en la reescritura de los textos, es decir, qué sugerencias fueron aceptadas e integradas en las nuevas versiones. Tanto las rúbricas diligenciadas como las versiones de los textos comentados se incluyeron en carpetas individuales en Google Drive.

Además, se llevaron a cabo dos grupos focales, grabados en video, para explorar las percepciones, reflexiones y aprendizajes de los participantes. El grupo focal es una técnica cualitativa de recolección de datos en el que un grupo de individuos seleccionados y reunidos por el investigador discuten y comentan sobre el tema objeto de estudio, de acuerdo con sus experiencias y percepciones (Rodas & Pacheco, 2020). El número de sujetos es variable, aunque suele situarse entre 6 y 14 (Pugsley, 2018). En este caso, ambos grupos focales incluyeron los 12 participantes de la experiencia y las coordinadoras asumieron el rol de moderadoras. El primer grupo focal se llevó a cabo al cierre de la sesión y se centró en la valoración de la experiencia de revisión. El segundo grupo focal se realizó al finalizar las sesiones del círculo de escritura, y buscó evaluar las percepciones sobre los impactos de la retroalimentación tanto en los textos como en los escritores. La Figura 1 sintetiza las técnicas de recolección:

Figura 1

Técnicas e insumos para la recolección de datos



Nota. Elaboración propia.

Análisis de datos

Para los datos de naturaleza cuantitativa se usaron técnicas de estadística descriptiva. En primer lugar, se empleó la medida de correlación intraclass que buscó establecer el grado de

concordancia entre las evaluaciones de ChatGPT y del par revisor. Esta medida calcula un análisis de varianza que compara la varianza entre casos e intracasos y da como resultado un coeficiente en un rango entre 0 y 1, el cero indica que no hay concordancia y el 1 que hay concordancia completa. Se usó el coeficiente de correlación intraclase de medida única, en el que se tomó una medida por cada fuente (evaluación de ChatGPT y evaluación del par) para cada texto y para cada criterio de la rúbrica. Se empleó el SPSS v.28 para calcular la consistencia en cada criterio de evaluación, así como un coeficiente global.

Por otra parte, a través de técnicas de distribución de frecuencias, se cuantificaron los comentarios incluidos en las rúbricas y en los textos de cada una de las fuentes. Posteriormente, se hizo una clasificación de los tipos de comentarios por propósito (descripción o sugerencia de modificación) y por contenidos o aspectos de la escritura en los que se centraban. Asimismo, se identificó la cantidad de cambios realizados en los borradores de las introducciones después de la retroalimentación.

Los datos de naturaleza cualitativa, recolectados a partir de las intervenciones de los participantes en los dos grupos focales, se procesaron con la técnica de análisis de contenido cualitativo apoyado en Atlas.Ti 23. El procedimiento incluyó: 1) la transcripción de los discursos orales de los participantes; 2) la codificación y etiquetado de los contenidos por afinidad temática; 3) la agrupación de la información en las categorías de pertinencia de la retroalimentación, calidad e impacto; y 4) la interpretación y sustento de los hallazgos a partir de ejemplos textuales.

Aspectos éticos de la investigación

Este trabajo no implicó riesgos físicos, emocionales ni sociales para los usuarios. Participaron voluntariamente mayores de edad, quienes aceptaron las condiciones de recolección de datos mediante un formato de consentimiento informado y quienes podían retirarse en cualquier momento del proceso. Se mantuvieron condiciones de confidencialidad y anonimato, asignando un código alfanumérico para identificar a los participantes. De igual modo, atendiendo a los posibles riesgos de uso de datos personales por parte de aplicaciones de inteligencia artificial, se decidió crear 12 usuarios genéricos y provisionales para que los estudiantes participaran en la experiencia de revisión.

Resultados

Los hallazgos se organizan en tres categorías, que buscan responder a las preguntas de investigación. En primer lugar, se analiza la congruencia entre las puntuaciones asignadas por los pares y por ChatGPT; en segundo lugar, se clasifican las sugerencias proporcionadas por ambas fuentes y su nivel de aceptación por parte de los escritores; finalmente, se da cuenta de las percepciones y valoraciones de los participantes respecto al proceso de retroalimentación.

Congruencia entre las fuentes de retroalimentación

Esta primera categoría busca identificar el grado de acuerdo entre las puntuaciones asignadas por ChatGPT y por los pares. Si bien las calificaciones no son el principal componente de una retroalimentación formativa, constituyen un punto de referencia para identificar los desempeños de los estudiantes. En este caso, la rúbrica de evaluación incluyó 10 criterios y una escala de calificación de 0 a 5.0, siendo 5 la puntuación máxima.

Como muestra la Tabla 1, los promedios de evaluación por cada criterio son similares, aunque las valoraciones de los pares son ligeramente mayores. Por otra parte, se encontró mayor variabilidad en las calificaciones de ChatGPT. Los aspectos con menor valoración por parte de ChatGPT fueron el planteamiento del objetivo de investigación, la cohesión, el estilo académico formal y la corrección gramatical y ortográfica.

Tabla 1

Puntuaciones de los textos

Criterio de evaluación	ChatGPT		Revisor par	
	Media	Desv. Est.	Media	Desv. Est.
Contextualización	4,56	0,15	4,6	0,17
Problema de investigación	3,99	0,45	4,21	0,28
Revisión de literatura	4,21	0,24	4,32	0,16
Objetivo general	3,58	0,40	4,06	0,25
Justificación	4,45	0,27	4,63	0,18
Coherencia	4,15	0,28	4,39	0,24
Cohesión	3,68	0,61	3,93	0,34
Estilo académico	3,64	0,46	4,13	0,44

Corrección gramatical y ortográfica	3,6	0,42	4,03	0,27
Formato de citación	4,25	0,32	4,06	0,2
Total	4,01	0,36	4,24	0,26

Nota. Elaboración propia con apoyo de SPSS v.28.

Para establecer la consistencia entre las evaluaciones, se calculó el coeficiente de correlación intraclase para cada criterio, en el que se encontraron medidas entre rangos aceptable y bueno. El coeficiente de correlación intraclase general fue de 0,870, con un p.valor < 0,005, lo que indica consistencia fiable entre las evaluaciones de las dos fuentes (Tabla 2).

Tabla 2

Coeficiente de correlación intraclase

	Correlación intraclase	Intervalo de confianza al 95 %		Prueba F con valor verdadero 0			
		Límite inferior	Límite superior	Valor	gl1	gl2	Sig
Medidas únicas	,770	,314	,938	7,708	9	9	,003
Medidas promedio	,870	,478	,968	7,708	9	9	,003

Nota. Elaboración propia con apoyo de SPSS v.28.

Efectividad de la retroalimentación

Esta categoría se centra en la identificación del impacto de la retroalimentación en los textos, es decir, si las sugerencias planteadas por ChatGPT y por los pares fueron incorporadas y aportaron al desarrollo de los borradores.

En este caso, se incluyen únicamente los comentarios que implicaban alguna acción por parte del escritor (sugerencia, recomendación), ya que tanto los pares revisores como ChatGPT incluyeron otros tipos de comentarios que no requerían ajustes (resaltar los avances, proporcionar información, expresar un punto de vista, etc.). La Tabla 3 muestra la frecuencia de los comentarios y de aceptación de ajustes para cada fuente de retroalimentación.

Tabla 3

Total de sugerencias y ajustes efectivos

Texto	# Palabras	ChatGPT		Revisor par	
		Sugerencias	Ajustes efectivos	Sugerencias	Ajustes efectivos
T1	2153	35	17 (49 %)	16	13 (81 %)
T2	2221	43	27 (63 %)	15	10 (67 %)
T3	1998	13	8 (62 %)	7	3 (43 %)
T4	2221	29	23 (79 %)	13	13 (100 %)
T5	2140	23	17 (74 %)	15	6 (40 %)
T6	2082	27	16 (59 %)	12	11 (85 %)
T7	1900	28	12 (43 %)	13	6 (67 %)
T8	2125	19	17 (89 %)	9	6 (67 %)
T9	2181	17	10 (59 %)	11	3 (27 %)
T10	2390	22	19 (86 %)	10	7 (70 %)
T11	2239	35	20 (57 %)	11	9 (82 %)
T12	2100	13	10 (69 %)	8	6 (75 %)
Total		304	195 (66 %)	140	85 (61 %)

Nota. Elaboración propia con apoyo de SPSS v.28.

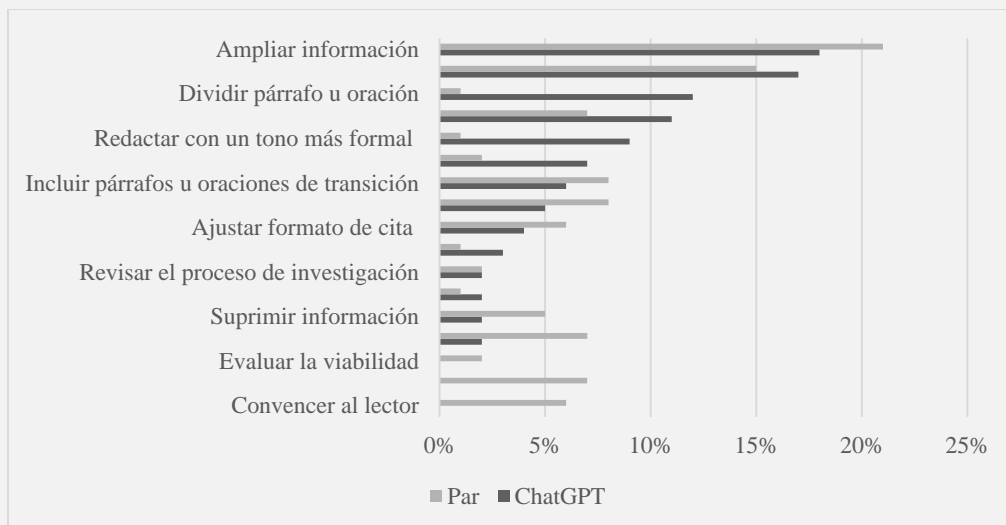
Como puede observarse, ChatGPT proporcionó en total 304 sugerencias; más del doble de las planteadas por los pares revisores. El promedio de sugerencias por texto de ChatGPT fue de 25,3, mientras que la media de los pares fue de 11,6. Es importante destacar que los participantes, con frecuencia, solicitaron a ChatGPT ampliaciones de sus comentarios, lo que pudo aumentar el número de recomendaciones.

El porcentaje de aceptación e incorporación de los ajustes por parte de los escritores es similar, 66 % para las sugerencias de ChatGPT y 61 % para las de los pares.

Aunque la retroalimentación se basó en un instrumento de evaluación con criterios y descriptores específicos, se evidenciaron diferentes énfasis en los tipos de sugerencias por cada fuente (Figura 2). Tanto ChatGPT como los pares revisores sugirieron en mayor medida ampliar información en el texto y cambiar conectores para mejorar la transición entre párrafos u oraciones.

Figura 2

Objeto de las sugerencias



Nota. Elaboración propia con apoyo de SPSS v.28.

Por su parte, ChatGPT presentó un mayor número de recomendaciones relacionadas con aspectos puntuales de la organización del texto, como se muestra en este comentario del texto 7, en el que se pide al participante que explicita el objetivo al cierre de la introducción:

“Evaluación: El objetivo de la investigación está implícito en la introducción y se refiere a la necesidad de indagar sobre los cambios en las prácticas de enseñanza debido a la virtualización de emergencia.

Sugerencia: Explicar de manera más explícita el objetivo de investigación al final de la introducción”.

De igual modo, esta IA se enfocó en aspectos microtextuales, como la segmentación de párrafos y oraciones, corrección ortográfica, ajustar el tono de la redacción a un estilo más académico y formal, cambiar signos de puntuación, etc. Por ejemplo, para el texto 4 sugiere cambiar las expresiones “se evidencia” por “se respalda” y “que señalan” por “quienes señalan”. Además, incorpora signos de puntuación y ajusta la tilde diacrítica de “cómo”:

“Aquí tienes algunos ejemplos específicos del texto proporcionado anteriormente que podrían ajustarse para cumplir un estilo académico formal:

1. Texto original:

‘...esta afirmación también se evidencia en investigaciones realizadas por Palos (2017), Marroquín y Valverde (2019) que señalan...’

Sugerencia:

‘...esta afirmación se respalda en investigaciones previas realizadas por autores como Palos (2017), Marroquín y Valverde (2019), quienes señalan...’

2. Texto original:

‘En este orden de ideas nos parece pertinente abordar esa relación y establecer como el docente puede llegar a mejorar sus prácticas a través del reconocimiento de esta interacción’.

Sugerencia:

‘En este contexto, consideramos relevante investigar esta relación y determinar cómo los docentes pueden mejorar sus prácticas mediante el reconocimiento de esta interacción’.

No obstante, en los casos en los que el escritor pedía orientaciones adicionales a la rúbrica, ChatGPT podía atender aspectos globales y proporcionar ideas para incluir en apartados, como es el caso de la justificación del texto 4:

“Enriquecer la justificación de tu investigación es crucial para proporcionar una base sólida y persuasiva. Aquí hay algunas ideas adicionales que podrías incluir:

1. Relevancia social ampliada:

Expande la discusión sobre la relevancia social de tu investigación, vinculándola con desafíos y necesidades actuales en la sociedad. ¿Cómo podría la mejora de las prácticas pedagógicas influir en aspectos más amplios de la educación y el desarrollo social?

2. Contribución al conocimiento existente:

Destaca cómo tu investigación se suma al conocimiento existente en el campo. ¿Qué vacíos o limitaciones específicas aborda tu estudio?

3. Impacto en la práctica docente:

Discute cómo los resultados de tu investigación podrían tener un impacto práctico en la enseñanza de las Ciencias Sociales. ¿Cómo podrían los hallazgos informar y mejorar las estrategias pedagógicas?”

En algunos casos, los escritores identificaron que al pedir al ChatGPT que detallara errores formales, este podía presentar “invenciones” de datos. En la valoración del texto 1, ChatGPT indica que hay aspectos que pueden mejorar en el criterio de corrección ortográfica y gramatical. Ante la solicitud de ampliación, muestra supuestos errores en el uso de mayúsculas en dos palabras, pero estas mayúsculas no aparecen en el texto original.

“Uso incorrecto de mayúsculas

- **Ejemplo:** ‘La institución *escuela* y sus maestros, están llamados a vehiculizar la lectura crítica como una de las principales herramientas de desarrollo del pensamiento crítico para comprender y apropiarse asertivamente de los artefactos de la cultura (Brunner, 2010; Coll, 1996)’.
- **Comentario:** No es necesario utilizar mayúsculas en ‘escuela’ y ‘están’. Las mayúsculas deberían reservarse para nombres propios o inicios de oraciones”.

Por su parte, los pares centraron la retroalimentación en aspectos más generales, relacionados con mantener la coherencia del texto mediante la reagrupación de información o la inclusión de párrafos de transición. En un comentario al texto 8 se recomienda una reorganización de las ideas:

“Esta parte de las pruebas PISA es importante porque muestra las deficiencias de los alumnos en lectura pero siento que es muy larga y es mejor acortarla un poco, pero, también se repite algo de esto en la justificación y queda cargado, creo que va solo en el problema y en la justificación más bien mostrar los beneficios que tendrían trabajar en lectura en primaria más allá de los resultados de PISA”.

Los pares, al compartir el contexto de producción con los escritores, hacen uso de sus experiencias y conocimientos sobre el proceso de investigación y proporcionan recomendaciones sobre aspectos como la situación comunicativa (ejemplo de sugerencia en texto 8) o la viabilidad de los trabajos (ejemplo en texto 3).

“Hay que mostrar más la importancia de hacer la investigación con estudiantes de primaria, de pronto mostrar alguna estadística de pruebas de lectura o algo que muestre que hay problemas con esa población, porque los datos que hay son de otros niveles”.

“Me inquieta la amplitud del problema, considerando que tenemos solo un año para terminar todo el proyecto y escribir el documento completo, no lo veo tan viable. Se podría

revisar si en el objetivo general es mejor analizar aspectos de la estrategia y no evaluar el impacto porque de pronto no alcanzas”.

Llama la atención que la mayoría de los textos presentaban errores en los formatos de citación, pero estos no fueron reconocidos por las dos fuentes de revisión. Se presentaron recomendaciones adicionales con muy baja frecuencia de aparición, estas se recogieron bajo la opción “Otro”.

Percepciones de los escritores

Esta categoría explora las percepciones de los participantes sobre la retroalimentación proporcionada por ChatGPT en comparación con los pares revisores. Se retoman las discusiones que se llevaron a cabo al cierre de la primera sesión de retroalimentación (grupo focal 1) y al finalizar el círculo de escritura (grupo focal 2). Los hallazgos se organizan a partir de tres criterios: la pertinencia o utilidad de la retroalimentación de ambas fuentes, la calidad de los comentarios y la valoración del impacto en el aprendizaje de la escritura.

En relación con la pertinencia, entendida como la utilidad del proceso de retroalimentación, se encontró una valoración positiva de los participantes, quienes destacaron los aportes de las revisiones a la transformación de sus textos y también a la configuración de su rol como escritores. Entre los aportes mencionan aprendizajes sobre la organización y estructuración de las ideas, delimitación del problema de investigación y reconocimiento de estrategias discursivas y estilísticas en el apartado introductorio:

“Muy importantes los comentarios de hoy, siento que me ayudan mucho a mejorar mi texto porque estaba un poco desorganizada mi cabeza y me voy con una estructura más clara que me permite concretar mi problema de investigación” (grupo focal 1, participante 3).

“Me parece que la rúbrica le da ese marco que se necesita para que los comentarios estén situados, respondan a lo que es, pero también sean amplios y permitan hacer discusiones y aportes al documento” (grupo focal 2, participante 9).

“Con la revisión tanto de humanos como de máquinas mejoré el planteamiento y toda la introducción (...) y aprendí cosas que no se suelen enseñar sobre las tesis...la mención a las citas, el lenguaje formal” (grupo focal 1, participante 2).

No se establecen grandes diferencias respecto a la utilidad de la retroalimentación entre las dos fuentes. Sin embargo, algunos participantes consideran que hay aspectos que ni ChatGPT ni los pares pueden apoyar en relación con la escritura, y que, por tanto, sería recomendable incluir también la revisión de un experto en lenguaje:

“Los dos son importantes y ayudan mucho, se deben mantener en las próximas revisiones, solo que ni la IA ni nosotros somos realmente expertos, tenemos errores... y eso lo puede disminuir un profesor de lenguaje” (grupo focal 1, participante 11).

Respecto a la calidad, los participantes tienen percepciones heterogéneas respecto a las características de los comentarios. En ese aspecto, algunos plantean que ChatGPT presenta sugerencias más detalladas y precisas que los compañeros, además de que queda un registro escrito de las consultas y preguntas adicionales, lo que ayuda a que los ajustes se incorporen de manera más efectiva en las nuevas versiones:

“El ChatGPT parece que se justifica más (...) si evalúa algo de manera negativa te da una sugerencia de cómo arreglarlo. Los compañeros pueden hacer comentarios más amplios sin justificarlos” (grupo focal 2, participante 3).

“Mientras que el ChatGPT te puede decir que cambies ‘ya que’ por ‘debido que’, una persona solo te anota ‘cambiar’” (grupo focal 2, participante 4).

“En la sesión anterior mi compañera me explicó tantas cosas, pero no la pude grabar y cuando fui a hacer la corrección ya no recordaba todo, por eso es mejor que quede por escrito como con el ChatGPT” (grupo focal 2, participante 7).

No obstante, otros participantes cuestionaron la calidad de la retroalimentación de ChatGPT, debido a que en ocasiones presentó “invenciones”. Estas aparecieron específicamente en los casos en que se pidió a la herramienta alguna precisión sobre errores en los aspectos formales de la escritura (ortografía o puntuación). Se valoró negativamente esta característica, en tanto dificultó el proceso de escritura de aquellas personas que no tienen mucha claridad sobre las normas gramaticales de la lengua:

Varias cosas que me marcó ni siquiera estaban en el texto, como mayúsculas o tildes que no existían y me decía que las quitara (...) Yo no soy experta en estos temas y tuve que tener mucho cuidado y eso me hizo tardar mucho más en decidir qué cambiaba y qué no y esto no optimizó mi proceso. (Grupo focal 2, participante 1)

Por otra parte, valoraron una visión más integrada de los pares revisores durante el proceso, quienes hicieron anotaciones sobre elementos extratextuales como el contexto de las investigaciones, estilos de redacción en textos de educación, viabilidad de las propuestas y procesos administrativos del programa académico. Esto se atribuye a que los pares están familiarizados tanto con los conocimientos como con los modos discursivos de la disciplina:

“Los compañeros están escribiendo y están viendo materias, entonces conocen más del tema, de la escritura, de lo que piden los asesores y de la vida en la escuela porque son maestros” (grupo focal 2, participante 6).

“Pude pensar más en si de verdad voy a poder terminar mi proyecto en el tiempo indicado y si de verdad me estoy ajustando a lo que nos piden acá” (grupo focal 2, participante 12).

Finalmente, se indagó con los participantes acerca del impacto de la retroalimentación en sus textos y en su aprendizaje de la escritura. Si bien se reiteraron las limitaciones sobre la calidad de ChatGPT y la importancia de incorporar una fuente experta adicional, los participantes percibieron mejoras a partir de ambas fuentes. Sobre los textos, destacaron aspectos como la puntuación y el estilo de redacción, modos discursivos específicos de la disciplina y articulación entre las ideas:

“[ChatGPT] Me ayuda mucho con la parte de la puntuación que a mí me cuesta mucho” (grupo focal 1, participante 2).

“Es muy bueno cuando [ChatGPT] hace la evaluación tipo lista de chequeo y te va diciendo si se cumplen o no los criterios, así estás tranquilo de que se incluyó todo lo necesario” (grupo focal 2, participante 5).

“Yo me confundía mucho con la primera o la tercera persona en la escritura, pero mis compañeros me ayudaron a definirme mejor y a ser congruente con la metodología de sistematización de experiencias que quiero hacer” (grupo focal, participante 6).

Pensé mucho en un comentario [de un par] sobre la manera en que suelen empezarse las introducciones en educación, que vienen desde afirmaciones muy generales, lo que choca un poco con mi estilo porque en ingeniería se arranca de una con el objetivo. (Grupo focal 2, participante 11)

Algunos indicaron cambios más amplios en sus procesos escriturales y en su motivación por la escritura, al reconocer la importancia del proceso de revisión y de escribir enfocándose en

el lector, así como las posibilidades que ofrecen múltiples herramientas digitales que apoyan la escritura académica:

Ahora en casi todo lo que escribo busco tener la mirada de otro lector, le he dicho a mi esposo que lea mi texto o pongo lectura en voz alta en Word e imagino que el texto no es mío para poder evaluarlo mejor. (Grupo focal 2, participante 6)

“No conocía muchas herramientas y empecé a usar el ChatGPT para planear, para ver cómo está mi documento, también he buscado otras como Spanish Checker para verificar, los verificadores antiplagio y demás” (grupo focal 2, participante 3).

“Me organizo mejor y como sé que me van a leer, así sea el ChatGPT, entonces escribo con más disciplina” (grupo focal 2, participante 7).

Discusión

Los hallazgos de esta investigación muestran un impacto positivo de la revisión entre pares y de la retroalimentación apoyada por IA generativa en el proceso de escritura en posgrados, así como similitudes entre ambas fuentes. Estos hallazgos coinciden con investigaciones anteriores, en tanto se identifica acuerdo entre las puntuaciones asignadas por la IA y por revisores humanos (Hackl et al., 2023; Mizumoto & Eguchi, 2023; Naismith et al., 2023), se reconoce el potencial de los LLM para detectar y abordar aspectos de nivel microtextual (Algarady et al., 2023; Tzirides et al., 2023; Yeşilyurt & Sezgin, 2023) y se valoran las ventajas de ChatGPT como tutor personalizado. Entre estas se encuentran la descripción detallada de las recomendaciones, mayor legibilidad de los comentarios y la posibilidad de interactuar para continuar ampliando la información, lo que ofrece más soporte para integrar los ajustes en nuevas versiones de los textos (Dai et al., 2023; Guo & Wang, 2024; Steiss et al., 2024).

Sin embargo, estos resultados también ponen en cuestión los alcances de ChatGPT en relación con herramientas de evaluación automatizada tradicionales, en tanto sus aportes se circunscriben, en su mayoría, a aspectos estructurales y formales. Si bien esto es interesante y complementa la labor de otros agentes, ChatGPT no está optimizado para fines de EAES y no cuenta con funciones como la anotación en el texto o subrayado, pero tampoco logra un análisis de dimensiones macrotextuales ni extratextuales.

Las imprecisiones y errores de ChatGPT son una de las principales limitaciones en el proceso de retroalimentación de la escritura. Aunque no son muy frecuentes, se presentan invenciones o “alucinaciones” con información dentro y fuera de los textos, lo cual representa desafíos para la escritura académica, en la que es necesaria una evaluación analítica que garantice la autenticidad y fiabilidad de los datos (Tzirides et al., 2023). Esta herramienta también carece de información actualizada que podría ser pertinente para poner en diálogo con los textos, aunque las nuevas versiones están trabajando en resolver este problema. La experiencia con el mundo, el compartir referentes, es justamente uno de los aspectos mejor valorados en relación con la retroalimentación proporcionada por los pares, quienes tienen una visión más integral y comprensiva de los textos. De ahí la importancia en el análisis a prácticas ligadas a la revisión como el discernimiento, la crítica frente a las respuestas de las IA generativas, la corroboración de la información y la implicación del escritor para mantener el rigor académico y la confiabilidad de su trabajo escrito (Guo & Lee, 2023).

En ese sentido, cabe recordar que la escritura de géneros como la tesis es un proceso complejo, que involucra distintos niveles que van más allá de la articulación coherente de palabras, oraciones y párrafos, que implica la generación de una voz, la apropiación de convenciones discursivas de las disciplinas, el vínculo con la realidad, entre otros aspectos (Lillis & Scott, 2015). Como plantea Cassany (2023):

Sería muy fácil escribir si fuera cierto, si bastaran un puñado de palabras y reglas de sintaxis y ortografía para escribir bien. Pero se necesita mucho más: tradiciones, retórica, estructuras discursivas, puntuación, originalidad (...) Escribir no es una regla, ni una técnica, ni una mecánica, ni un algoritmo. Es un proceso largo, complejo y estratégico: hay que hallar a la audiencia, descubrir sus necesidades, darse cuenta de lo que sabe y de lo que ignora. Hay que conocer la tradición escrita de cada ámbito para poder respetarla o saltársela con la máxima cortesía. (pp. 30-31)

Identificar los alcances de las herramientas es esencial para que los usuarios puedan determinar de manera crítica y ética los usos que harán de ellas. Aun así, hay características latentes de ChatGPT que podrían favorecer los procesos escriturales y complementar la labor de los revisores humanos. Entre ellas se encuentra que ChatGPT está disponible para el escritor en cualquier momento, proporciona retroalimentación inmediata y tiene múltiples formas de

interacción con el usuario (Steiss et al., 2024; Tajik & Tajik, 2024), sin contar que requiere de un mínimo entrenamiento y bajo costo económico. Con marcos epistemológicos y prácticos adecuados, en el mediano plazo quizá pueda constituirse en un sistema de tutoría inteligente para la escritura, con orientaciones personalizadas y ajustadas a las necesidades de cada escritor, sin que esto implique que se reemplacen las labores de los escritores ni de los revisores humanos (Carlson et al., 2023). Cabe destacar, además, que esta investigación se llevó a cabo menos de un año después de la masificación del acceso a ChatGPT y que se están experimentando cambios vertiginosos en relación con las IA generativas, que logran cada vez más efectividad en sus resultados.

Los dispositivos didácticos de acompañamiento a la escritura, entre ellos los círculos de escritura en los posgrados, deben continuar explorando sobre el rol que pueden tener las IA generativas en estos espacios. En ese sentido, aunque circunscrita a un contexto particular, esta investigación aporta evidencia empírica sobre cómo se pueden integrar aplicaciones de IA generativa en procesos de formación en escritura académica en espacios horizontales y de gestión autónoma del aprendizaje como los círculos de escritura, sin desconocer las responsabilidades institucionales y la necesidad de acompañamiento de asesores expertos. Comprender y evaluar los impactos de este tipo de experiencias redundan en la toma de decisiones pedagógicas más razonadas y conscientes.

Se requieren estudios que exploren las diversas funcionalidades de las herramientas de la IA y su impacto en la productividad, en la calidad de los textos y en la generación de experiencias de aprendizaje a través de la escritura. La integración de la IA en la escritura de la tesis y en los procesos de retroalimentación ratifica la idea de que no se trata una actividad única, sino de diferentes actividades desafiantes para el escritor como analizar, sintetizar, criticar, argumentar. Probablemente con el apoyo de la IA y de la articulación de esta a los planes de estudio se podrían hacer aportes a la formación en escritura en investigación.

Este artículo responde a la necesidad de literatura sobre escritura de tesis (Aitchison & Guerin, 2014) y la integración de herramientas generadas por IA en el proceso. La literatura existente explora los problemas de escritura de tesis (Odena & Burgess, 2017), la falta de instrucciones sobre cómo escribir (Graham, 2018) y lo referido al razonamiento y la argumentación (Wentzel, 2018), pero no necesariamente analiza la IA como asistente de la

producción de la tesis o como apoyo en los procesos de retroalimentación de una manera que contribuya como experiencia de aprendizaje.

Así, este estudio destaca el papel significativo de ChatGPT en la producción textual. Su versatilidad en las fases de revisión y edición muestra su potencial para refinar las metodologías de enseñanza de la escritura académica. Sin embargo, los desafíos relacionados con la adaptación, la ética, el acceso y la preparación de los docentes deben abordarse para aprovechar al máximo sus beneficios en los procesos de enseñanza y revisión entre pares (Escalante et al., 2023; Tziriades et al., 2023).

Entre las limitaciones investigativas se sitúan la amplitud de la muestra y el contexto particular de la experiencia de trabajo, debido a que se contó con la participación voluntaria de 12 estudiantes que estaban interesados en avanzar en la escritura de su tesis y que ya venían explorando herramientas de IA generativa. El grupo reducido posibilitó, además, un contacto permanente con las coordinadoras y un seguimiento minucioso de las actividades de retroalimentación entre pares y retroalimentación con ChatGPT, aspectos que generalmente no se encuentran en aulas regulares de escritura o en las asesorías de tesis.

En líneas futuras de investigación sería recomendable ampliar el número de participantes y de textos, así como explorar otras dimensiones del impacto de la retroalimentación en la escritura, que trasciendan la incorporación de las recomendaciones y que analicen la manera como los escritores las comprenden, evalúan y asumen. La perspectiva de los participantes, sus representaciones y expectativas sobre las IA generativas son tan importantes como las orientaciones pedagógicas (Dai et al., 2023; Escalante et al., 2023; Steiss et al., 2024). Del mismo modo, se podrían incluir mediciones en distintos momentos y probar con modelos y rúbricas centrados en aspectos más globales de los textos. Debido a que el proceso de escritura es complejo e implica múltiples factores, valdría la pena explorar el impacto de la IA generativa en otras fases como la generación de ideas, la planificación de los textos o durante la textualización.

Encontramos en esta investigación nuevas líneas de trabajo en los círculos de escritura, transferibles a otros dispositivos de acompañamiento como las tutorías en centros de escritura o las retroalimentaciones en cursos de escritura, espacios en los que pueden aprovechar las IA generativas con propósitos específicos de aprendizaje. Se debe considerar, por supuesto, la necesidad de desarrollar un pensamiento crítico por parte de los actores involucrados y de propiciar

reflexiones sobre los alcances y limitaciones de estas herramientas (Rob, 2023; Sun & Hoelscher, 2023).

Conclusiones

Esta investigación se propuso evaluar los efectos de la retroalimentación proporcionada por ChatGPT y por revisores pares a la escritura del apartado introductorio de tesis de posgrado en educación en el marco de un círculo de escritura. Al respecto, se identificó que ambas fuentes hacen aportes relevantes tanto a la mejora del texto como al proceso de los escritores.

Se encontró que las puntuaciones asignadas por ChatGPT y por pares son congruentes, aunque ligeramente más altas por parte de los pares. ChatGPT proporcionó un poco más del doble de sugerencias de ajuste que los revisores humanos y los escritores incluyeron más de la mitad de las recomendaciones de ambas fuentes. Los pares aportaron a aspectos globales como la coherencia, la progresión temática y la viabilidad investigativa, mientras que ChatGPT aportó a elementos más específicos de orden estructural y formal, evidenciando así una complementariedad entre ambas fuentes evaluativas.

Los 12 participantes valoraron positivamente la pertinencia, calidad e impacto de las retroalimentaciones en el avance de sus textos y en sus procesos escriturales en general, aunque también identificaron limitaciones que tendrán que considerarse en el marco de los dispositivos de acompañamiento a la escritura que incorporen IA generativa en los procesos de retroalimentación.

Aunque se reconocen las limitaciones relacionadas con la muestra y las particularidades de contexto de investigación, los hallazgos sugieren oportunidades de articulación entre distintas fuentes de retroalimentación para aportar a la mejora de géneros especializados como las tesis, así como a los procesos de aprendizaje de los escritores, lo que permitiría una mayor transferencia de los conocimientos y competencias desarrolladas. A partir de este estudio se identifican fortalezas que podrían considerarse en propuestas similares, especialmente en círculos de escritura y en programas de tutorías, entre ellas la necesidad de establecer una estrategia de acompañamiento deliberada para la escritura en posgrados, en la que se discutan y practiquen distintas formas de evaluación formativa que permitan avanzar en la escritura, a través de múltiples fuentes de retroalimentación, como es el caso de los profesores, asesores, pares y herramientas automatizadas.

Es fundamental, en este caso, reconocer la naturaleza de los aportes de cada una de las fuentes y la necesidad de asumir posturas críticas como escritores. El uso de las herramientas debe estar mediada por el reconocimiento de sus alcances y limitaciones, la construcción efectiva y la evaluación iterativa de *prompts*, el modelado mediante ejemplos y el establecimiento de formas de interacción que permitan precisar, ampliar o reformular la información de acuerdo con las necesidades de cada escritor.

Si bien las aplicaciones de IA generativa están todavía en desarrollo y no sustituyen la acción de los evaluadores humanos, se constituyen en una opción interesante, efectiva y de bajo costo para apoyar la personalización de los aprendizajes de la escritura en espacios educativos horizontales como los círculos de escritura. Para ello, se requiere de una formación reflexiva y crítica de todos sus participantes en relación con el uso de las IA generativas y demás tecnologías de apoyo a la escritura.

Referencias

- Abreu, J. (2015). Síndrome Todo Menos Tesis (TMT). All But Thesis Syndrome. *Daena: International Journal of Good Conscience*, 10(2), 246-259. [http://www.spentamexico.org/v10-n2/A14.10\(2\)246-259.pdf](http://www.spentamexico.org/v10-n2/A14.10(2)246-259.pdf)
- Aitchison, C. (2009). Research Writing Groups [Grupos de redacción de Investigaciones]. In J. Higgs, D. Horsfall, & S. Grace (Eds.), *Writing Qualitative Research on Practice* (pp. 253-263). Sense.
- Aitchison, C., & Guerin, C. (2014). *Writing Groups for Doctoral Education and Beyond: Innovations in Practice and Theory* [Grupos de redacción para la educación doctoral y más allá: innovaciones en la práctica y la teoría]. Routledge.
- Algaraady, J., & Mahyoob, M. (2023). ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners [Capacidades de ChatGPT para detectar y analizar errores de escritura experimentados por estudiantes de inglés como lengua extranjera]. *Arab World English Journal (AWEJ) Special Issue on CALL*, (9), 3-17. <https://dx.doi.org/10.24093/awej/call9.1>

- Atlas, S. (2023). ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI [ChatGPT para la educación superior y el desarrollo profesional: una guía para la IA conversacional]. *College of Business Faculty Publications*. https://digitalcommons.uri.edu/cba_facpubs/548
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? [Fuentes de retroalimentación en la redacción de ensayos: ¿retroalimentación generada por pares o por IA?]. *International Journal of Educational Technology in Higher Education*, 21(1), 1-15. <https://doi.org/10.1186/s41239-024-00455-4>
- Barton, D., & Hamilton, M. (2004). La literacidad entendida como práctica social. En V. Zavala, M. Niño-Murcia, & P. Ames (Eds.), *Escritura y sociedad. Nuevas perspectivas teóricas y etnográficas* (pp. 109-139). Red para el desarrollo de las ciencias sociales en el Perú.
- Barton, D., & Lee, C. (2013). *Language Online: Investigating Digital Texts and Practices [Lenguaje en línea: Investigación de textos y practicas digitales]*. Routledge.
- Biermann, O. C., Ma, N. F., & Yoon, D. (2022). From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies [De herramienta a compañero: los guionistas quieren que los escritores de IA respeten sus valores personales y sus estrategias de escritura]. *Designing Interactive Systems Conference*. <https://dl.acm.org/doi/10.1145/3532106.3533506>
- Carlson, M., Pack, A., & Escalante, J. (2023). Utilizing OpenAI's GPT-4 for written feedback [Utilizando GPT-4 de OpenAI para comentarios escritos]. *TESOL Journal*, 15(2), 1-7. <https://doi.org/10.1002/tesj.759>
- Cassany, D. (2023). *Metáforas sospechosas. Charlas mestizas sobre la escritura*. Anagrama.
- Castillo-González, W. (2023). The importance of human supervision in the use of ChatGPT as a support tool in scientific writing [La importancia de la supervisión humana en el uso de ChatGPT como herramienta de apoyo en la escritura científica]. *Metaverse Basic and Applied Research*, 2(29). <https://doi.org/10.56294/mr202329>
- Castro, Y. (2024). Factores relacionados al estado Todo-Menos-Tesis en la formación de grado de odontólogos de una universidad peruana. *Revista de Investigación e Innovación en Ciencias de la Salud*, 6(2), 204-218. <https://doi.org/10.46634/riics.291>

- Colombo, L. M., Iglesias, A., Kiler, M., & Saez, V. (2022). Grupos de escritura en el postgrado: experiencias de tesistas: Experiencias de tesistas. *Espacios en blanco, Revista de Educación, 1*(32), 163-172. <https://doi.org/10.37177/UNICEN/EB32-322>
- Creswell, J., & Plano, V. L. (2018). *Designing and conducting mixed methods research* [Diseñar y realizar investigaciones con métodos mixtos]. Sage Publications.
- Cuevas-Solar, D., & Arancibia, B. (2020). Percepciones y expectativas de docentes de ingeniería y educación en torno a la retroalimentación en tareas de escritura. *Formación universitaria, 13*(4), 31-44. <https://dx.doi.org/10.4067/S0718-50062020000400031>
- Cuthbert, D., Spark, C., & Burke, E. (2009). Disciplining writing: The case for multi-disciplinary writing groups to support writing for publication by higher degree by research candidates in the humanities, arts, and social sciences [Disciplinar la escritura: el caso de los grupos de escritura multidisciplinarios para apoyar la escritura para publicación de grado superior por parte de candidatos a investigación en humanidades, artes y ciencias sociales.]. *Higher Education Research & Development, 28*(2), 137-149. <https://doi.org/10.1080/07294360902725025>
- Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y., Gasevic, D., & Chen, G. (2023). *Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT* [¿Pueden los modelos de lenguaje grandes proporcionar retroalimentación a los estudiantes? Un estudio de caso sobre ChatGPT]. <https://doi.org/10.35542/osf.io/hcgzj>
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing [De la escritura humana al texto generado por inteligencia artificial: examinando las perspectivas y amenazas potenciales de ChatGPT en la escritura académica]. *Biology of Sport, 40*(2), 615-622. <https://doi.org/10.5114/biol sport.2023.125623>
- Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence [Sistemas automatizados de evaluación de escritura: una revisión sistemática de Grammarly, Pigai y Criterion con una perspectiva sobre las

- direcciones futuras en la era de la inteligencia artificial generative]. *Education and Information Technology*, 29, 14151-14203. <https://doi.org/10.1007/s10639-023-12402-3>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference [Comentarios generados por IA sobre la escritura: información sobre la eficacia y la preferencia de los estudiantes de ENL]. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00425-2>
- Espeche, M., Colombo, L., & Rodas, E. (2024). Un grupo de escritura virtual doctoral: factores que influyen en su funcionamiento según los participantes. *Zona Próxima*, (40), 43-65. <https://doi.org/10.14482/zp.40.211.968>
- Farazouli, A., Cerratto, T., Bolander, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices [¡Hola GPT! ¿Adiós examen domiciliario? Un estudio exploratorio del impacto de los chatbots de IA en las prácticas de evaluación de los docentes universitarios]. *Assessment and Evaluation in Higher Education*, 1-13. <https://doi.org/10.1080/02602938.2023.2241676>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research [Un análisis FODA de ChatGPT: implicaciones para la práctica y la investigación educativas]. *Innovations in Education and Teaching International*, 1-15. <https://doi.org/10.1080/14703297.2023.2195846>
- Godoy, A., & Damián Ferro, M. (2021). Baja graduación en posgrados: aplicabilidad y validez de los conceptos y categorías de Pierre Bourdieu para el análisis e interpretación de una carrera de Maestría. *Acta Sociológica*, (82), 105-133. <https://doi.org/10.22201/fcpys.24484938e.2020.82.79453>
- Gómez, M., & Alzate, M. (2018). Persistencia en los estudios de doctorado: revisión de literatura. *Voces y Silencios. Revista Latinoamericana de Educación*, 9(2), 107-125. <https://doi.org/10.18175/vys9.2.2018.07>

- Graham, S. (2018). Changing how writing is taught [Cambiando la forma en que se enseña la escritura]. *Review of Research in Education*, 43(1). <https://doi.org/10.3102/0091732X18821125>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing [¿Resistirlo o abrazarlo? Examinando el potencial de ChatGPT para respaldar la retroalimentación de los docentes en la escritura de inglés como lengua extranjera]. *Educ Inf Technol*, 29, 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for Enhancing Critical Thinking Skills. *Journal of Chemical Education*, 100(12), 4876-4883. <https://doi.org/10.1021/acs.jchemed.3c00505>
- Hackl, V., Müller, A. E., Granitzer, M., & Sailer, M. (2023). Is GPT-4 a reliable rater? Evaluating Consistency in GPT-4 Text Ratings. *ArXiv*. <http://arxiv.org/abs/2308.02575>
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays [Una comparación a gran escala de ensayos escritos por humanos versus ensayos generados por ChatGPT]. *Scientific Reports*, 13(1), 1-11. <https://doi.org/10.1038/s41598-023-45644-9>
- Hill-Yardin, E., Hutchinson, M., Laycock, R., & Spencer, S. (2023). A Chat (GPT) about the future of scientific publishing [Una charla (GPT) sobre el futuro de la publicación científica]. *Brain Behav Immun*, (110), 152-154. <https://doi.org/10.1016/j.bbi.2023.02.022>
- Hsiao, Y.-P., Klijn, N., & Chiu, M.-S. (2023). Developing a framework to re-design writing assignment assessment for the era of Large Language Models [Desarrollar un marco para rediseñar la evaluación de tareas de escritura para la era de los modelos de lenguaje grandes]. *Learning Research and Practice*, 9(2), 148-158. <https://doi.org/10.1080/23735082.2023.2257234>
- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature [Análisis del papel de ChatGPT como asistente de redacción en el nivel de educación superior: una revisión sistemática de la literatura]. *Contemporary Educational Technology*, 15(4), Article e464. <https://doi.org/10.30935/cedtech/13605>

- Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning [Explorando el papel de la inteligencia artificial para facilitar la evaluación del desempeño en escritura en el aprendizaje de una segunda lengua]. *Languages*, 8(4), 1-16. <https://doi.org/10.3390/languages8040247>
- Lillis, T., & Scott, M. (2015). Defining academic literacies research: issues of epistemology, ideology and strategy [Definición de la investigación sobre alfabetizaciones académicas: cuestiones de epistemología, ideología y estrategia]. *Journal of Applied Linguistics and Professional Practice*, 4(1), 5-32. <https://doi.org/10.1558/japl.v4i1.5>
- Lillo-Fuentes, F., Venegas, R., & Lobos, I. (2023). Evaluación automatizada y semiautomatizada de la calidad de textos escritos: una revisión sistemática. *Perspectiva Educacional*, 62(2), Artículo 1420. <https://dx.doi.org/10.4151/07189729-Vol.62-Iss.2-Art.1420>
- Lozić, E., & Štular, B. (2023). Fluent but not factual: A comparative analysis of ChatGPT and other AI chatbots' proficiency and originality in scientific writing for humanities [Fluido pero no objetivo: un análisis comparativo de la competencia y originalidad de ChatGPT y otros chatbots de IA en la escritura científica para humanidades]. *Future Internet*, 15(10), 1-26. <https://ideas.repec.org/a/gam/jftint/v15y2023i10p336-d1259029.html>
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., & Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges [ChatGPT y grandes modelos lingüísticos en el mundo académico: oportunidades y desafíos]. *BioData Mining*, 16(20). <https://doi.org/10.1186/s13040-023-00339-9>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring [Explorando el potencial del uso de un modelo de lenguaje de IA para la puntuación automatizada de Ensayos]. *Research Methods in Applied Linguistics*, 2(2), Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Molina, V., & López-Gil, K. (2020). Estado de la cuestión de los centros y programas de escritura en Latinoamérica. *Revista Colombiana de Educación*, 78, 97-119. <http://doi.org/10.17227/rce.num78-8066>

- Naismith, B., Mulcaire, P., & Burstein, J. (2023). Automated evaluation of written discourse coherence using GPT-4 [Evaluación automatizada de la coherencia del discurso escrito utilizando GPT-4]. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 394-403). Association for Computational Linguistics.
- Narciss, S., & Zumbach, J. (2020). *Formative Assessment and Feedback Strategies* [Evaluación formativa y estrategias de retroalimentación]. Springer International Publishing.
- Nazar, R., & Renau, I. (2023). Estilector: un sistema de evaluación automática de la escritura académica en castellano. *Perspectiva Educativa*, 62(2), 37-59. <https://dx.doi.org/10.4151/07189729-vol.62-iss.2-art.1427>
- Núñez, J. A., & da Cunha, I. (2022). El impacto del uso de herramientas automáticas de ayuda a la redacción en el proceso de escritura de estudiantes universitarios. *Círculo de Lingüística Aplicada a la Comunicación*, 89, 131-144. <https://doi.org/10.5209/clac.73906>
- Odena, O., & Burgess, H. (2017). How doctoral students and graduates describe facilitating experiences and strategies for their thesis writing learning process: A qualitative approach [Cómo estudiantes y graduados de doctorado describen experiencias y estrategias facilitadoras para su proceso de aprendizaje en la redacción de tesis: un enfoque cualitativo]. *Studies in Higher Education*, 42(3), 572-590. <https://doi.org/10.1080/03075079.2015.1063598>
- OpenAI. (2023). *GTP-4 Research*. <https://openai.com/research/gpt-4>
- Pugsley, L. (2018). Focus groups, young people and sex education [Grupos focales, jóvenes y educación sexual]. In *Gender and Qualitative Research* (pp. 114-130). Routledge.
- Rob, R. (2023). *ChatGPT in Higher Education: Artificial Intelligence and its Pedagogical Value* [ChatGPT en la educación superior: la inteligencia artificial y su valor pedagógico]. Pressbook.
- Rodas, F., & Pacheco, P. (2020). Grupos focales: marco de referencia para su implementación. *INNOVA Research Journal*, 5(3), 182-195. <https://doi.org/10.33890/innova.v5.n3.2020.1401>

- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing [Comparación de la calidad de los comentarios humanos y ChatGPT sobre la escritura de los estudiantes]. *Learning and Instruction, 91*, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Street, B. (2017). New Literacies, New Times: Developments in Literacy Studies [Nuevas literacidades, nuevos tiempos: Avances en los Estudios de Literacidad]. In B. V. Street & S. May (Eds.), *Literacies and Language Education* (pp. 3-15). Springer International Publishing.
- Strobl, C., Ilhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies [Soporte digital para la escritura académica: una revisión de tecnologías y pedagogías]. *Computers & Education, 131*, 33-48. <https://doi.org/10.1016/j.compedu.2018.12.005>
- Sun, G. H., & Hoelscher, S. H. (2023). The ChatGPT storm and what faculty can do [La tormenta ChatGPT y lo que los profesores pueden hacer]. *Nurse Educator, 48*(3), 119-124. <https://doi.org/10.1097/nne.0000000000001390>
- Tajik, E., & Tajik, F. (2024). A Comprehensive Examination of the Potential Application of Chat GPT in Higher Education Institution [Un examen exhaustivo de la posible aplicación de Chat GPT en instituciones de educación superior]. *Social Science Research Network*. <http://dx.doi.org/10.2139/ssrn.4699304>
- Tambunan, A. R. S., Andayani, W., Sari, W. S., & Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback [Investigar los problemas lingüísticos de los estudiantes de inglés como lengua extranjera utilizando Grammarly como retroalimentación de evaluación de escritura automatizada]. *Indonesian Journal of Applied Linguistics, 12*(1), 16-27. <https://doi.org/10.17509/IJAL.V12I1.46428>
- Tzirides, A., Saini, A., Zapata, G., Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., Abrantes, R., Whiting, J., & Polyxeni, N. (2023). Generative AI: Implications and Applications for Education [IA generativa: implicaciones y aplicaciones para la educación]. *arXiv*. <https://arxiv.org/pdf/2305.07605>

- Wentzel, A. (2018). *A guide to argumentative research writing and thinking: Overcoming challenges* [Una guía para la redacción y el pensamiento de investigaciones argumentativas: superando desafíos]. Routledge.
- Yeşilyurt, Y. E., & Sezgin, S. (2023). Automated Writing Evaluation with a Large Pre-Trained Language Model a Preliminary Study [Evaluación de escritura automatizada con un gran modelo de lenguaje previamente entrenado: un estudio preliminar]. *SSRN*. <http://dx.doi.org/10.2139/ssrn.4545244>
- Zhai, N., & Ma, X. (2023). The Effectiveness of Automated Writing Evaluation on Writing Quality: A Meta-Analysis [La eficacia de la evaluación automatizada de la escritura sobre la calidad de la escritura: un metanálisis]. *Journal of Educational Computing Research*, 61(4), 875-900. <https://doi.org/10.1177/07356331221127300>