

Uso de Inteligencia Artificial Generativa para retroalimentar escritura académica en procesos de Formación Inicial Docente

Providing academic writing feedback assisted by Generative Artificial Intelligence in initial teacher education contexts

Carlos Ossa*  ; Carlos Willatt 

Facultad de Educación. Universidad Autónoma de Chile, Chile.

Resumen

Este artículo presenta el trabajo de implementación de la Inteligencia Artificial Generativa (IAG) para proporcionar retroalimentación descriptiva a 578 respuestas a casos pedagógicos elaboradas por una cohorte de estudiantes ($n = 347$) de 7 programas de formación inicial docente en una Facultad de Educación en Chile. Para retroalimentar dichas respuestas, se diseñó un proceso denominado RECaP-GPT, que considera la acción humana y utiliza ChatGPT-4 como herramienta de apoyo docente. La retroalimentación entregada se basa en criterios preestablecidos y conocidos por los estudiantes. La retroalimentación producida por la IAG fue validada y proporcionada a los estudiantes para mejorar sus evidencias de aprendizaje a lo largo de un semestre de clases. Se discute la capacidad de apoyo que esta herramienta posee para retroalimentar de manera formativa y a gran escala el trabajo escrito de los estudiantes en un corto período de tiempo. Además, se plantean limitaciones observadas en el proceso de implementación que requieren una mayor investigación.

Palabras clave: inteligencia artificial generativa; estudio de casos; retroalimentación descriptiva; escritura académica; formación inicial docente;

* Autor para correspondencia: carlos.ossa@uaautonoma.cl

Abstract

This paper presents the implementation of Generative Artificial Intelligence (GAI) for providing descriptive feedback on 578 responses to pedagogical cases written by a cohort of students (n = 347) from seven initial teacher education programs at a Faculty of Education in Chile. To deliver feedback on these responses, a process called RECaP-GPT was developed, which incorporates human action and utilizes ChatGPT-4 as a teaching support tool. The feedback provided is based on pre-established criteria known to the students. The feedback generated by the GAI was validated and delivered to the students to enhance their evidence of learning throughout a semester of classes. The paper discusses the capacity of this tool to offer formative and extensive feedback on students' written work in a short period of time. Furthermore, it raises limitations observed in the implementation process that necessitate further research.

Keywords: generative artificial intelligence; case study; descriptive feedback; academic writing; initial teacher education;

INTRODUCCIÓN

La retroalimentación como práctica evaluativa es fundamental para la mejora en los procesos de aprendizaje de todos los niveles educativos, al proveer a los estudiantes información relevante para comprender y evaluar su propio desempeño, además de ejercer un indudable efecto en su sentido de autoeficacia (Sutton, 2012; Wiggins, 2012). Con esta práctica se busca que el estudiante, con el apoyo de un docente experto, pueda realizar los ajustes necesarios para acortar la brecha entre su nivel de desempeño actual, y el nivel que se espera que alcance en relación a un determinado criterio, objetivo o meta formativa (Sadler, 1989; Shute, 2008).

En general, cuando los docentes retroalimentan la evidencia de aprendizaje de los estudiantes, deben considerar, además de su calidad, que la retroalimentación sea entregada al estudiante en un tiempo que le permita obtener provecho de tal ejercicio (Sutton, 1998).

Quien provea retroalimentación debe cautelar al menos tres aspectos que no son obvios. Wiliam (1999, 2011) ha enfatizado que:

- Quienes han recibido una calificación sin haber recibido retroalimentación previa, no presentan mejoras en sus desempeños posteriores.
- El mensaje formativo de una retroalimentación se anula si va acompañado con una calificación sumativa, haciendo que el trabajo del docente sea prácticamente inútil (Wiliam, 2011).
- Cuando la calificación de un desempeño se posterga hasta el final, quienes durante el proceso reciben una retroalimentación clara, descriptiva y propositiva, mejoran sus desempeños hasta en un 30% en promedio.

La rúbrica es el instrumento que, basado en la selección de criterios de evaluación relevantes para la tarea (Shepard, 2006), permite caracterizar la calidad de los desempeños de los estudiantes, describiendo un estándar de desempeño máximo por alcanzar y niveles de desempeño intermedio que considera los errores que típicamente registran quienes están en un proceso de aprendizaje. La identificación de tales criterios de evaluación y la descripción de esos desempeños, permiten enfocar tanto el trabajo del docente como de los estudiantes hacia aspectos que se estiman que son relevantes de demostrar.

Estas consideraciones hacen que la labor de retroalimentación en su bidireccionalidad, sea demandante tanto desde la perspectiva del trabajo de diseño de instrumentos y de su activación por parte de los docentes, como la del trabajo y compromiso de los estudiantes con sus propios procesos de aprendizaje (Carless & Winstone, 2020; Henderson et al., 2019). Esto adquiere complejidades propias cuando se busca retroalimentar la escritura académica en la educación superior, con propósitos formativos y de manera descriptiva (Tunstall & Gipps, 1996), proveyendo retroalimentación a gran escala y en tiempo acotado (Nedjat-Haiem & Cooke 2021; Svanes & Andersson-Bakken, 2023).

Frente a este desafío, desde hace más de 50 años se ha planteado como una solución a este problema el desarrollo de sistemas de retroalimentación automatizados para apoyar el trabajo docente (Deeva et al., 2021). En la década del 60 del siglo pasado se diseñaron e implementaron los primeros sistemas automatizados para la calificación de ensayos, los que siguen siendo perfeccionados en la actualidad (Doewes & Pechenizkiy, 2021; Ke & Ng, 2019).

El desarrollo de sistemas o herramientas automatizadas que, más allá de la calificación, puedan retroalimentar de manera descriptiva el desempeño de los estudiantes al producir textos escritos con respuestas abiertas es, sin embargo, aún muy incipiente. Los sistemas y plataformas basadas en Inteligencia Artificial Generativa (IAG), particularmente los Transformadores Generativos Preentrenados (GPT), ofrecen algunas alternativas cuyos resultados pueden ser prometedores (Dai et al., 2023). Dentro de estos últimos, destaca el *chatbot* ChatGPT, desarrollado por la empresa estadounidense OpenAI, liberado al uso público global el 30 de noviembre de 2022. A pesar de su corta vida, ChatGPT ha ganado rápidamente notoriedad en todo el mundo debido a su accesibilidad y versatilidad, habiendo ya evidencia de uso cada vez más frecuente en el ámbito educativo (Galindo-Domínguez et al., 2023; Gazquez et al., 2023; UNESCO, 2023).

1.1 La escritura de análisis de casos pedagógicos en la formación inicial docente

La escritura de análisis de casos como recurso de preparación y ensayo para la toma de decisiones en situaciones de complejidad profesional es ampliamente conocido tanto en la formación inicial como continua de muchas profesiones (Hare & Portelli, 2003),

particularmente en el estudio del derecho, de la medicina y la pedagogía, aunque también en el estudio de los eventos de accidentes de aviación en el contexto de la formación de pilotos aéreos comerciales. La necesidad de estudiar errores que pueden ser previstos por medio del análisis y simulación de situaciones profesionales complejas es una preocupación fundamental para quienes se desempeñan con el factor humano como objeto de trabajo.

Por medio del estudio de casos se abre un espacio para que esas situaciones emergentes que necesitan una toma de decisiones informada sean previstas. Esto permite la proposición de diversas alternativas de solución, sabiendo que la retroalimentación para esa tarea, preparada y provista por un experto, pueda dar orientación efectiva respecto de las causas y consecuencias de las acciones decididas en el diseño de resolución de un caso (Berliner, 2001).

En el caso de la pedagogía, el estudio de casos es una estrategia que permite el aprendizaje y el ejercicio de habilidades y competencias asociadas a la escritura académica y, por medio de ella, la integración de la teoría y la práctica docente en torno a situaciones pedagógicas, así como la reflexión y la toma de decisiones fundada frente a problemas y dilemas éticos que surgen en el desempeño de esta profesión (Martínez Sánchez, 1999; Schön, 1991).

Un caso pedagógico que corresponde a una situación verosímil de la práctica, compleja en la medida de las experiencias vitales de docentes que sí las han experimentado, permite al aprendiz poder poner en ejercicio las posibilidades y alternativas de solución a problemas propuestos y a proponer la toma de decisiones profesionales, con el mayor fundamento teórico y práctico disponible al momento de la actividad.

La toma de postura reflexiva que exige un caso pedagógico impone, a quien lo busca resolver, la tarea de decidir fundadamente. Esa decisión, que puede ser acertada o no, no dejará de ser formadora de autoconciencia (*self consciousness*) profesional. El análisis de las decisiones propuestas como respuesta al caso, intenciona la mirada crítica hacia el interior del estudiante y propicia la búsqueda de la definición de la identidad profesional y del sentido de sus acciones profesionales (Hare y Portelli, 2003).

El presente artículo muestra el trabajo de implementación de IAG para retroalimentar 587 respuestas escritas a casos pedagógicos, elaboradas por una cohorte de estudiantes (n = 347) de programas de pedagogía de modalidad concurrente en una Facultad de Educación en Chile. La retroalimentación entregada contempla la descripción de la calidad de escritura provista por el estudiante, su adscripción al tipo de escritura académica aceptada para la tarea, su organización sintáctica, el uso de referentes teóricos y legales, así como algunos aspectos formales básicos.

A continuación, se presenta el método utilizado para la implementación de la IAG. Luego, se exponen algunos hallazgos relevantes vinculados al modelo de retroalimentación implementado. Finalmente, se discuten los alcances de esta investigación y algunas proyecciones y limitaciones asociadas al ámbito de la formación inicial docente.

MÉTODO

2.1 Participantes

La implementación de la IAG para retroalimentar la escritura académica de casos pedagógicos se realizó en el contexto de una asignatura de 18 semanas de duración, en el semestre Primavera de 2023, impartido a 347 estudiantes provenientes de siete programas distintos de pedagogía de modalidad concurrente en pregrado, y que cursaban su sexto semestre de estudios, de un total de ocho. El propósito de la asignatura era ahondar en los fundamentos éticos de la profesión docente para la resolución de situaciones y dilemas pedagógicos, más allá de las particularidades del área de especialización disciplinar de los programas de estudio.

Para abarcar la totalidad de estudiantes, la cohorte de estudiantes fue dividida en 10 secciones a cargo del mismo número de docentes tanto de la planta académica regular como adjunta y con distinta formación disciplinar. Durante las 18 semanas de duración del curso, los docentes sostuvieron 6 reuniones de coordinación conducidas por los autores.

2.2 Procedimiento

Dado el fuerte componente ético de la asignatura, una de las tareas fundamentales trabajadas con los estudiantes consistió en la resolución de 6 casos pedagógicos, de los cuales 4 fueron retroalimentados en su totalidad con ayuda de IAG, el último de ellos calificado. Estos casos, encomendados periódicamente a partir de la cuarta semana de clases, consistían en una situación pedagógica no asociada al tratamiento disciplinar de los contenidos, para la que los estudiantes debían proponer una respuesta haciendo referencia explícita a saberes pedagógicos, la normativa curricular vigente y teorías del aprendizaje y el desarrollo humano estudiadas anteriormente en el curso de los respectivos programas académicos.

Antes de trabajar en la resolución de los casos, cada uno de los docentes presentó y discutió la rúbrica de criterios preestablecidos con los estudiantes de su sección para que éstos comprendieran desde el principio cómo serían evaluadas sus respuestas.

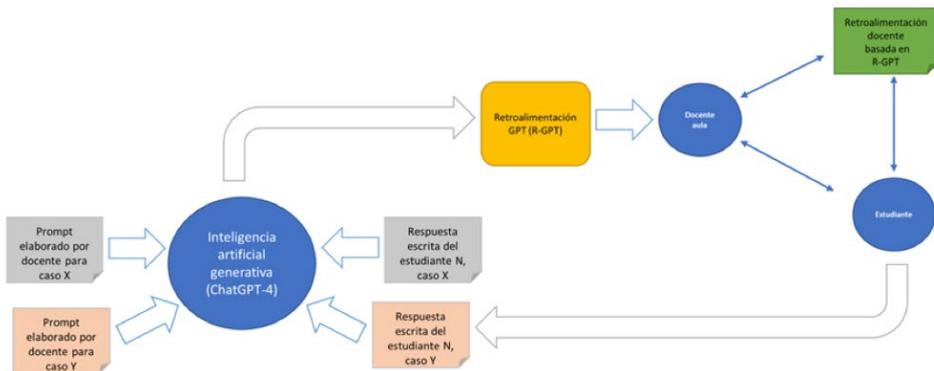
Para elaborar sus respuestas a los casos pedagógicos, los estudiantes de todas las secciones del curso debían completar una plantilla de respuesta en formato *Microsoft Word* diseñada especialmente para ese propósito. Tanto la descarga de la plantilla como el envío del documento con la respuesta fueron efectuadas mediante el sistema de gestión de aprendizajes *Canvas* asociado al curso y sus secciones. Para este estudio se consideraron las 587 respuestas de estudiantes retroalimentadas con ayuda de IAG.

Al término del curso, se solicitó a los estudiantes que, de manera voluntaria y anónima, escribieran un breve comentario sobre su experiencia al recibir retroalimentación generada por IAG.

2.3 Instrumentos

Para retroalimentar las respuestas de los estudiantes se modeló un proceso denominado *Retroalimentación de Escritura de Casos Pedagógicos basada en Transformador Generativo Preentrenado* (RECaP-GPT). Este proceso implica por un lado la incorporación activa de la acción humana, tanto del docente como del estudiante, y el uso de una plataforma de IAG cargada con la API (*Application Programming Interface*) de ChatGPT-4, lanzada por OpenAI en julio de 2023, como herramienta de apoyo docente (Figura 1).

Figura 1. Modelo Retroalimentación de Escritura de Casos Pedagógicos, basado en Transformador Generativo Preentrenado (RECaP-GPT).



Para elaborar un *prompt* básico que instruyera a la IAG del tipo de retroalimentación esperada para cada caso pedagógico, los autores tomaron como referencia dos planteamientos teóricos relativos a la evaluación del aprendizaje y al componente lingüístico de la actividad: la Evaluación para el Aprendizaje (*Assessment for Learning*), centrada en la evaluación de desempeños académicos por medio de retroalimentación de tipo descriptiva que permite la mejora de esos desempeños (Black & Wiliam, 1998; Torrance & Pryor, 1998; Stobart, 2014; Tunstall & Gipps, 1996); y una aplicación al discurso inspirada en la teoría de *chunks* y del lenguaje formulaico (Chomsky, 1965; Wray, 2002) para proponer la “silueta” de una estructura discursiva mínima para la respuesta a los casos pedagógicos.

A partir de los modelos teóricos de referencia se desarrolló una rúbrica de carácter descriptivo con tres criterios. Estos criterios permiten distinguir aspectos relevantes para la retroalimentación de la calidad de un texto producido por un estudiante como análisis de una situación pedagógica, la que denominamos genéricamente como *caso* (Ossa & Willatt, 2023).

Con cada caso se ajustaba el *prompt* de acuerdo a los requerimientos de la tarea propuesta, de tal forma que la IAG entregara también su propuesta de retroalimentación con los fundamentos requeridos para argumentarla.

Los criterios de evaluación usados en el *prompt* son:

1. *Silüeta de respuesta a un caso*. Este criterio implica que el texto escrito por el estudiante posee una estructura básica de tres partes, y un cuarto aspecto relacionado con la extensión de la respuesta:

- a) Replanteamiento del problema que presenta el caso y la instrucción que se da a partir de él. El estudiante reescribe el problema planteado en el caso.
- b) Análisis y justificación de la propuesta. El estudiante utiliza en su escritura los referentes epistemológicos, teóricos, legales, etc., que permiten sustentar los análisis necesarios para su explicación o justificación.
- c) Diseño de actividad y propuesta de evaluación o seguimiento para resolver el problema y verificar su efectividad. Considera los recursos pedagógicos necesarios para su solución con pertinencia o efectividad. El estudiante provee un diseño completo de una actividad que busca resolver el problema.
- d) Extensión de la respuesta. El texto de la respuesta al caso planteado no debe ser inferior a 500 palabras ni superior a 600 palabras.

2. *Calidad de la propuesta de resolución del caso*. De acuerdo a este criterio, el estudiante describe una estrategia de resolución del problema del caso, analizando el problema planteado, proponiendo intervenciones docentes basadas en el contexto del caso y reconociendo los recursos pedagógicos disponibles, fundamentando teóricamente tales estrategias y actividades en base a las teorías del desarrollo, del aprendizaje, o la legislación asociada a la profesión docente en Chile.

3. *Calidad de la comunicación escrita*. Este criterio examina si el texto manifiesta un foco expositivo que, siguiendo una organización lógica, maneja las normas de expresión escrita y sus recursos de coherencia y cohesión. Asimismo, se espera que el texto se atenga a las convenciones ortográficas, usando un registro y tono adecuado para la situación comunicativa.

2.4 Análisis de datos

Las respuestas de los estudiantes recibidas en *Canvas* fueron ingresadas en el recuadro de chat de la plataforma de IAG, la que retroalimentó cada respuesta individualmente de acuerdo al *prompt* correspondiente al caso pedagógico. Como fue dicho, en el *prompt* se estableció la estructura del texto de retroalimentación de salida que la plataforma de IAG debía entregar.

Dependiendo de la calidad de la respuesta del estudiante, la plataforma de IAG tardaba entre 45 y 90 segundos en retroalimentar completamente una respuesta. En este tiempo de procesamiento la plataforma calificaba la respuesta del estudiante con los conceptos entregados en la rúbrica (Completo/Incompleto; Muy Bueno, Bueno, Suficiente, Insuficiente) y describía la calidad del texto presentado, indicando las mejoras que el estudiante debía realizar para alcanzar mejores desempeños de acuerdo a lo descrito en la rúbrica.

Con los criterios de evaluación preestablecidos mostrados más arriba y que están descritos con mayor especificidad en otro lugar (Ossa & Willatt, 2023), la plataforma de IAG elaboró, por cada una de las 587 respuestas de estudiantes, un texto de retroalimentación como el que se expone en la siguiente tabla.

Tabla 1. Ejemplo de retroalimentación generada por IAG, validada por docente

EVALUACIÓN GENERAL		
CRITERIO 1 - Silueta de respuesta a un caso		
Parte 1 (Replanteamiento):	Completo	
Parte 2 (Análisis y justificación):	Incompleto	El análisis y la justificación de la intervención son aceptables, pero podrían ser más robustas al involucrar de manera más explícita teorías de aprendizaje y desarrollo. Por ejemplo, podría mencionar la teoría del aprendizaje social de Bandura para justificar el enfoque de aprendizaje entre pares.
Parte 3 (Diseño):	Completo	
CRITERIO 1.1 - Extensión de la respuesta:	Completo	

EVALUACIÓN GENERAL

CRITERIO 2 - Calidad de la propuesta

Parte 1 (Estrategia):	Nivel Bueno	La estrategia que propone el estudiante es pertinente, pero podría beneficiarse de una elaboración más detallada y una conexión más clara con las teorías del aprendizaje y desarrollo.
Parte 2 (Fundamento):	Nivel Suficiente	Aunque se hace referencia a la teoría sociocultural de Vygotsky, sería beneficioso proporcionar una mayor elocución de cómo esta teoría y posiblemente otras teorías de aprendizaje y desarrollo respaldan específicamente las actividades propuestas.

CRITERIO 3 - Comunicación Escrita

Parte 1 (Foco):	Nivel Muy Bueno	
Parte 2 (Escritura)	Nivel Bueno	Si bien la mayor parte del texto está bien redactada, existen algunos errores ortográficos y de puntuación que podrían distraer al lector y dificultar la comprensión.
Parte 3 (Tono)	Nivel Muy Bueno	

CONCLUSIONES

El estudiante muestra una clara comprensión del problema planteado y desarrolla una propuesta de intervención lógica y detallada. Sin embargo, la respuesta podría beneficiarse de una justificación teórica más robusta y de mejoras ortográficas. En futuras propuestas, el estudiante podría beneficiarse de:

- Revisar y corregir errores ortográficos y de puntuación.
 - Hacer explícito cómo las actividades propuestas están respaldadas por las teorías de aprendizaje y desarrollo relevantes.
 - Proporcionar un análisis más profundo de los problemas a abordar y cómo se espera que -las actividades propuestas mejoren estos problemas.
-

Antes de entregar al estudiante su retroalimentación, cada docente de sección validó el contenido de la propuesta de retroalimentación en dos niveles. Por un lado, el docente debía verificar:

- que no se hubiera producido algún efecto de *alucinación* de ChatGPT, por ejemplo, calificar como *Aceptable* algo que se había conceptualizado como *Bueno*;
- que no hubiera uso de neologismos, de expresiones en inglés o anglicismos creados por la IAG, intercalados en el castellano requerido;
- la correcta discriminación por parte de ChatGPT de aspectos por mejorar del contenido del texto escrito por el estudiante.

El otro nivel de validación estaba dado por la participación del docente en la verificación de que los fundamentos teóricos, legales o de saberes docentes usados por el estudiante, correspondieran efectivamente a lo requerido en la construcción de un argumento legítimo para abordar el caso pedagógico. Ante un error conceptual el docente devolvía la retroalimentación propuesta a la IAG para generar una nueva versión más completa, que incluía el aspecto que debía mejorar de esa retroalimentación.

Al inicio del curso, la necesidad de validar las propuestas de retroalimentación provistas por la IAG fueron más frecuentes que hacia el final, cuando ésta se adecuó progresivamente a las indicaciones del *prompt* y a las sucesivas correcciones que se le dio. La retroalimentación validada se entregaba a cada estudiante a través de *Canvas*. En clase presencial se propiciaba el análisis de las respuestas, de sus aciertos o dificultades, a veces en relación docente-estudiantes, otras veces estudiante-estudiante, con la rúbrica a la vista.

RESULTADOS

Luego de la validación docente realizada a las propuestas de retroalimentación por el modelo RECaP-GPT, se proveyó retroalimentación descriptiva a 587 respuestas escritas por estudiantes, en base a los criterios de evaluación. La retroalimentación basada en IAG resultó ser eficiente en varios sentidos:

- ChatGPT tardó hasta 90 segundos en generar un texto base (ver Tabla 1) para la retroalimentación descriptiva de un texto de entre 500 a 600 palabras escrito por un estudiante.
- Todos los estudiantes recibieron por *Canvas* su retroalimentación validada en al menos tres días hábiles, y no después de 10 días hábiles después de haber entregado su trabajo.
- El tiempo que requirió la validación de la retroalimentación provista por la IAG fue considerablemente menor al que de manera regular toma validar externamente una retroalimentación de esas características hecha sin este tipo de apoyo.

Hay consenso entre los autores de que un efecto virtuoso de una retroalimentación continua y enfocada en el logro de aprendizajes preestablecidos es la mejora de la calidad del desempeño, cuando el estudiante adopta las sugerencias indicadas en una retroalimentación de estas características (Wiggins, 2012; Wiliam, 2012). Los resultados obtenidos en este estudio no permiten conclusiones definitivas a este respecto. Sin embargo, los comentarios que estudiantes de esta asignatura han compartido con sus docentes, escritos al término del semestre, pueden servir como antecedente preliminar del uso provechoso de la IAG para una mejora de la escritura académica, pues la percepción general de los estudiantes fue positiva.

Estudiantes de distintas secciones aseguran que “la retroalimentación proporcionada por la inteligencia artificial es excepcionalmente precisa y personalizada, permitiendo a nosotros como estudiantes recibir orientación adaptada a necesidades individuales.” (E1).

Un segundo aspecto destacado por los estudiantes es que la retroalimentación entregada es “clara y me ayudó a darme cuenta de mis errores generales. Se da un *feedback* personalizado, lo que se aprecia mucho sabiendo el tiempo que toma hacer esto estudiante por estudiante” (E2), destacando, por un lado, que las evidencias evaluadas pueden ser mejoradas con más trabajo del estudiante, y aunque sean retroalimentaciones individualizadas, los errores detectados refieren a la evidencia de aprendizaje presentada y no a la persona que la presentó. Por otro lado, si este estudiante percibió la cantidad de trabajo que significa retroalimentar de buena forma grandes cantidades de texto en tiempo breve, señala que la calidad de la retroalimentación recibida fue buena. Este comentario cobra mayor relevancia al provenir de futuros docentes, ya que la retroalimentación como estrategia de evaluación constituye un elemento central del trabajo profesional para el cual se están formando, particularmente en el contexto chileno, que se ajusta a los requerimientos de evaluación formativa y retroalimentación que establece el Decreto 67/2018.

Un tercer aspecto resaltado por los estudiantes es que la retroalimentación recibida indicó alternativas de solución a argumentos incompletos o mal referenciados. Ante el uso de una alternativa argumental equivocada, la IAG propuso el uso de un sustento teórico distinto al usado por el estudiante. Un estudiante señala que “mi retroalimentación fue muy completa, [...] agradezco que ante la elección errónea de una teoría, la corrección [retroalimentación] brinde otra a modo de ejemplo para saber cuál sería más adecuado usar y por qué” (E3). En este caso, la retroalimentación incluso pudo modelar una manera de enfrentarse a futuras situaciones análogas, señalando que “es muy útil al momento de hacer nuestras propias correcciones. Muy completas todas las razones de por qué se asignó la calificación y rango de respuesta.” (E3)

DISCUSIÓN Y CONCLUSIONES

A la luz del proceso de implementación del modelo RECaP-GPT diseñado especialmente para una asignatura impartida en una Facultad de Educación en Chile, es posible afirmar que la IAG es una herramienta altamente eficiente para apoyar la labor docente de retroalimentación. La IAG puede ejecutar la tarea de retroalimentación en tiempos brevísimos y con gran nivel de detalle, atendiendo las especificidades de cada respuesta entregada por los estudiantes. Sin duda, este rendimiento tecnológico es excepcional y abre nuevas posibilidades para la retroalimentación a gran escala de la escritura escolar y académica en general, y de la escritura de análisis de casos pedagógicos en particular. Esto último es crucial para apoyar los procesos de formación inicial docente de cara al desafío de profesionalizar a futuros docentes, en tiempos en que existe una gran demanda de docentes en los distintos niveles educativos a nivel nacional e internacional.

Otro resultado de la implementación de la IAG para retroalimentar respuestas a casos pedagógicos fue el modelamiento de la acción profesional docente enfocada en el logro de los aprendizajes preestablecidos respecto del desarrollo de las acciones pedagógicas relacionadas a ellos.

El uso de una plataforma de IAG posibilita la retroalimentación de un mayor número de ejercicios de casos pedagógicos y, por lo tanto, permite a los estudiantes abordar más frecuentemente situaciones que exigen una toma de postura profesional. Este modelamiento permitió ampliar en los estudiantes ámbitos de apropiación del discurso pedagógico que no eran previsibles al inicio de la asignatura. Por ejemplo, el uso cada vez más frecuente de la referencia al trabajo de autores relevantes de teorías del aprendizaje, del desarrollo humano (UNESCO, 2021), o la referencia a la legalidad vigente que rige la profesión docente en Chile, fueron evidentes. Esos contenidos, que habían sido estudiados sin mayor aplicación práctica en semestres anteriores, fueron utilizados y rescatados, de acuerdo a un estudiante:

“Antes no sabía bien cómo aplicar los conocimientos que sabía, lo teórico es difícil de entender en la vida real. El entender cómo responder a un caso me ayudó a sacar del papel mis conocimientos y ponerlos en práctica. Desde antes sentía una responsabilidad como profesor, pero me gusta sentir que ahora está guiada por una ética, la cual puedo respaldar con teorías acertadas. Me gusta cómo me siento apto para ser profesor en cualquier sala. Sé ahora aún más cuál es la importancia de observar una situación antes de actuar.” (E4)

Para facilitar una redefinición de la comprensión de la relación entre teoría y praxis, como sugiere el estudiante, es preciso ejercitarse en el estudio de casos, observar muchas y distintas situaciones. Pero si se hacen más ejercicios como estos, es preciso poder proveer retroalimentación útil en un tiempo razonable y con cierto detalle.

El entusiasmo y las perspectivas promisorias que pudieran surgir de la implementación de la IAG en contextos educativos, sin embargo, no deberían conducir a la conclusión apresurada acerca de la posibilidad de automatizar completamente los procesos de retroalimentación. Por esta razón, conviene señalar algunas de las limitaciones derivadas del proceso de implementación de la IAG en este estudio. Sin pretensión de completitud, se han identificado al menos tres niveles de limitaciones: técnicas, pedagógicas y éticas.

La primera limitación técnica asociada al modelo de ChatGPT-4 utilizado en este estudio tiene que ver con el acceso al mismo. A diferencia de la versión 3.5 de ChatGPT, que aún es accesible gratuitamente previo registro, ChatGPT-4 requiere la compra permanente de crédito o un abono mensual que es cuantificado a través de *tokens*. Los *tokens* se van utilizando según la extensión y el tipo de texto de entrada y salida. Sin desconocer el legítimo afán de lucro que las empresas desarrolladoras tengan, en este caso OpenAI, los costos de operación de una plataforma de IAG pueden ser un gran desincentivo para su implementación y masificación en contextos educativos, especialmente en países como Chile donde las instituciones de educación tienen un dispar acceso a recursos públicos y privados.

En cuanto a la experiencia de uso del modelo ChatGPT-4, cabe señalar que éste no ejecuta de buena forma el conteo de palabras y en muchos casos no detecta errores ortográficos, aspectos que bien podrían considerarse dentro de una rúbrica descriptiva para evaluar la calidad de la escritura. Asimismo, frente a un mismo texto de entrada suele generar retroalimentaciones disímiles e incluso alucinaciones.

Por estos motivos, no sólo conviene esperar nuevas versiones de ChatGPT, sino también tener claridad que ChatGPT es sólo uno de los posibles modelos de IAG disponibles en la actualidad. En este sentido, es necesario seguir explorando otras plataformas de IAG que tal vez pueden tener un mejor rendimiento para los propósitos de la retroalimentación aquí expuesta.

Una limitación pedagógica evidente es la tentación de excluir la figura del docente de aula del ciclo pedagógico en general, y de la retroalimentación evaluativa en particular, delegando el proceso completo a la IAG. A este respecto, cabe señalar que la determinación de los criterios de evaluación posee una dimensión normativa y se manifiesta a nivel de aula. El docente humano debe establecer qué es lo importante, valioso y pertinente que se aprenda. Asimismo, los niveles de calidad manifestados en estándares esperados debieran siempre estar cruzados con la condición de educabilidad de los estudiantes (López & Tedesco, 2002).

Otra limitación de este orden es la manifiesta y extendida dificultad en estudiantes de todos los niveles educativos de trabajar autónomamente en la mejora de sus evidencias de aprendizaje, si no hay una calificación asociada. La respuesta negativa ante la pregunta si la actividad es con nota tiende a disminuir la posibilidad de fijar un impacto de la retroalimentación formativa.

Por último, una limitación de orden ético refiere a que a la tradicional caja negra que disputa el concepto de mente, se suma la caja negra de la IAG, en este caso ChatGPT. Aunque las lógicas de lingüística informática que están detrás de ChatGPT puedan ser hasta cierto punto conocidas, se desconoce exactamente cómo funcionan sus algoritmos. Al reportar saberes específicos, por ejemplo, por qué escoger entre un autor u otro para argumentar, la IAG no indica en su retroalimentación de dónde ni cómo obtiene la información que sirve para construir ese *feedback*, o si esa información es de dominio público o de otro tipo. En el mercado de la información digitalizada, es sabida la preeminencia de uso de determinados antecedentes convertidos en algoritmos que están sometidos a oferta y demanda.

Paradójicamente, aunque puedan aparecer oportunidades asociadas, por ejemplo, la sugerencia de incluir una referencia a una teoría o un autor no contemplado inicialmente para construir un argumento, existe simultáneamente el riesgo de limitar el espectro de recursos que permitan el pensamiento y reflexión humana.

Todos estos aspectos requieren mayor desarrollo en investigaciones posteriores en las que habría que analizar los resultados del desempeño de esta retroalimentación, contrastar con la perspectiva de estudiantes y de docentes y observar con entusiasmo y cautela, al mismo tiempo, los avances de nuevas herramientas de IAG que esperan ser usadas.

REFERENCIAS

- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482 [https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Black, P., & Wiliam, D. (1998). *Inside the black box. Raising standards through classroom assessment*. King's College London.
- Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28, 150-163. <https://doi.org/10.1080/13562517.2020.1782372>
- Chomsky, N. (1965). *Aspects of theory of syntax*. MIT Press.
- 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. <https://doi.org/10.35542/osf.io/hcgzj>
- Decreto 67 de 2018. Aprueba normas mínimas nacionales sobre evaluación, calificación y promoción y deroga los decretos exentos n° 511 de 1997, n° 112 de 1999 y n° 83 de 2001, todos del Ministerio de Educación. <https://www.bcn.cl/leychile/navegar?idNorma=1127255>

- Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerd, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162, 1-43. <https://doi.org/10.1016/j.compedu.2020.104094>
- Doewes, A. & Pechenizkiy, M. (2021). On the limitations of human-computer agreement in automated essay scoring. En *Proceedings of The 14th International Conference on Educational Data Mining*, 475-480. https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_243.pdf
- Galindo-Domínguez, H., Delgado, N., Losada, D. & Etxabe, J. M. (2023). An analysis of the use of artificial intelligence in education in Spain: The in-service teacher's perspective. *Journal of Digital Learning in Teacher Education*. <https://doi.org/10.1080/02602938.2021.1902467>
- Gazquez Linares, J. J., Pérez Fuentes, M. & Suazo Galdames, I. (2023). Aprovechando el Potencial de la Inteligencia Artificial en la Educación: Equilibrando Beneficios y Riesgos. *European Journal of Education and Psychology*, 16(1), 1-8. <https://doi.org/10.32457/ejep.v16i1.2205>
- Hare, W., & Portelli, J-P. (2003). *What to do? Case studies for teacher educators*. EdPhil Books.
- Henderson, M., Ryan, T., Boud, D., Dawson, P., Phillips, M., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research and Development*, 38(7), 1401-1416. <https://doi.org/10.1080/07294360.2019.1657807>
- Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. En *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6300-6308. <https://www.ijcai.org/proceedings/2019/0879.pdf>
- López, N. & Tedesco, J. (2002). *Las condiciones de educabilidad de los niños y adolescentes en América Latina: documento para discusión*. UNESCO IIEP.
- Martínez Sánchez, A. (1999). El estudio de casos como técnica didáctica. Estudio de una experiencia en la enseñanza universitaria. *Innovación Educativa*, 9, 25-53.
- Nedjat-Haiem, M. & Cooke, J. (2021). Student strategies when taking open-ended test questions. *Cogent Education*, 8(1), 1-22. <https://doi.org/10.1080/2331186X.2021.1877905>
- Ossa, C. & Willatt, C. (2023). Retroalimentación efectiva basada en inteligencia artificial generativa. Criterios de evaluación para la escritura de casos pedagógicos. <http://doi.org/10.32457/12728/102782023120>

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Schön, D. (1991). *The reflective turn: Case studies in and on educational practice*. Teachers College Press.
- Shepard, L. (2006). *La evaluación en el aula*. Instituto Nacional para la Evaluación de la Educación.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189
- Stobart, G. (2014). *The expert learner*. Open University Press.
- Sutton, R. (1998). School wide assessment: improving teaching and learning. New Zealand Council for Educational Research.
- Sutton, P. (2012). Conceptualizing feedback literacy: knowing, being, and acting. *Innovations in Education and Teaching International*, 48(1), 31-40.
- Svanes, I. K. & Andersson-Bakken, E. (2023). Teachers' use of open questions: investigating the various functions of open questions as a mediating tool in early literacy education. *Education Inquiry*, 14(2), 231-250. <https://doi.org/10.1080/20004508.2021.1985247>
- Torrance, H. & Pryor, J. (1998). *Investigating formative assessment. Teaching, learning and assessment in the classroom*. Open University Press.
- Tunstall, P. & Gipps, C. (1996). Teacher feedback to young children in formative assessment. *British Educational Research Journal*, 22(4), 389-404.
- UNESCO (2021). *Formadores de docentes en seis países de América Latina. Instituciones, prácticas y visiones*. UNESCO/UDP.
- UNESCO (2023). *ChatGPT and artificial intelligence in higher education: quick start guide*. <https://unesdoc.unesco.org/ark:/48223/pf0000385146>
- Wiggins, G. (2012). Seven keys to effective feedback. *Feedback for Learning*, 70(1), 10-16.
- Wiliam, D. (1999). Formative assessment in mathematics. *Equals: Mathematics and Special Education Needs*, 5(3) 8-11)
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Wiliam, D. (2012). Feedback: part of a system. *Feedback for learning*. 70(1), 30-34
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Recibido: 29 de octubre 2023
Aceptado: 20 de diciembre 2023