

1

# “Die Hausklausur”: Cuando la IA también se examina

“The Take-Home Exam”: When AI Is Also Put to the Test



AUTORA

**Birgit Strotmann<sup>1</sup>**

[birgit.strotmann@comillas.edu](mailto:birgit.strotmann@comillas.edu)  <https://orcid.org/0000-0002-8869-2341>

<sup>1</sup>Departamento de Traducción e Interpretación y Comunicación Multilingüe, Universidad Pontificia Comillas.



PALABRAS CLAVE | KEYWORDS

Inteligencia Artificial Generativa (IAG); Aprendizaje de idiomas; Examen a libro abierto;  
Pensamiento crítico; Innovación en la evaluación

Generative Artificial Intelligence; Language learning; Open-book exam; Critical thinking;  
Assessment innovation



## RESUMEN

Este trabajo surge como respuesta al impacto que el uso generalizado de la inteligencia artificial generativa (IAG) ha tenido en la enseñanza de lenguas extranjeras, especialmente en el desarrollo del trabajo autónomo del estudiante. Las actuales herramientas de IAG permiten la producción de textos de alta calidad, aunque a menudo sin un aprendizaje real asociado. Para abordar este reto, se propuso al alumnado de la asignatura *Alemán B2.1* realizar una de las pruebas objetivas en formato de “examen a libro abierto” (*Hausklausur*), con la instrucción expresa de utilizar cualquier herramienta de IAG disponible. El examen se diseñó para fomentar la reflexión crítica sobre las respuestas generadas por la IA, con un doble objetivo: promover el aprendizaje de alemán mediante la evaluación, verificación y corrección activa de dichas respuestas, y, en segundo lugar, generar conciencia sobre las posibilidades y las limitaciones de estas tecnologías. Como parte de la intervención, se aplicó un breve cuestionario anónimo para valorar su impacto. Los resultados mostraron una mayor conciencia crítica respecto a la IAG, aunque el alumnado no manifestó una percepción clara de haber aprendido más alemán mediante esta metodología. De cara al futuro, se recomienda integrar de forma sistemática tareas que impliquen el uso consciente de la IAG, perfeccionar el diseño de los exámenes a libro abierto y adoptar una metodología más robusta para medir el impacto, como un enfoque pre-post que permita comparar resultados de aprendizaje y actitudes antes y después de la intervención.

## ABSTRACT

This paper responds to the significant impact that the widespread use of generative artificial intelligence (GAI) has had on foreign language education, particularly on students' autonomous learning. Current GAI tools enable the production of high-quality texts, yet often without the corresponding impact on language acquisition on the part of the student. To address this challenge, students in the German B2.1 course were asked to complete one of their midterm exams in a “take-home exam” (*Hausklausur*) format, with explicit instructions to use any available GAI tools. The exam was designed to encourage critical reflection on AI-generated responses, with two main objectives: first, to promote German language learning through the evaluation, verification, and correction of AI outputs; and second, to foster awareness of the affordances and limitations of these technologies. As part of the intervention, a brief anonymous questionnaire was administered to assess its impact. The results indicated an increased critical awareness regarding GAI, although students did not report a clear perception of improved German learning through this method. Looking ahead, it is recommended to systematically integrate tasks that require conscious use of GAI, refine the design of open-book exams, and adopt a more robust methodology to measure impact — such as a pre-post design that compares learning outcomes and attitudes before and after the intervention.



## 1. INTRODUCCIÓN

La irrupción de la inteligencia artificial generativa (IAG) en la educación superior ha supuesto un cambio paradigmático, cuestionando principios fundamentales como la autoría, la originalidad, los procesos de aprendizaje y los sistemas de evaluación. Al mismo tiempo, ha abierto nuevas oportunidades, entre ellas la personalización del aprendizaje, la (semi) automatización de ciertas tareas docentes y la consiguiente liberación de tiempo del profesor para otras actividades. En consecuencia, como McDonald et al. (2025) señalan, muchas instituciones de educación superior han incorporado la IAG en sus políticas, currículos y programas, aunque todavía falta investigación sobre los cambios requeridos en las prácticas docentes y sobre los efectos reales en los resultados de aprendizaje de los estudiantes.

Aunque muchas instituciones aún se encuentran en proceso de adaptación, el estudiantado ha ido un paso por delante del profesorado (Fundación CYD, 2025), adoptando rápidamente la IAG como un recurso imprescindible en su vida académica. Diversos estudios constatan un uso generalizado de la IAG: más del 90 % del estudiantado la emplea habitualmente para tareas como redacción, revisión y traducción de textos, apoyo en la investigación, preparación de exámenes y análisis de datos (Higher Education Policy Institute, 2025; Jisc, 2025). Sin embargo, este uso se acompaña de preocupaciones éticas y pedagógicas. Según Mirea et al. (2025) y Niño-Carrasco et al. (2025), los estudiantes manifiestan inquietud por una posible dependencia, la falta de fiabilidad de los resultados y una disminución de su capacidad crítica, al tiempo que demandan formación práctica sobre redacción de *prompts*, uso ético de la IAG y políticas institucionales (JISC, 2025).

El impacto de la IAG en el aprendizaje depende, en gran medida, del enfoque que adopten los estudiantes. Pallant et al. (2025), en un estudio con 192 participantes, evidencian que un uso orientado al dominio del conocimiento ("mastery approach") se asocia con mejores resultados que un uso meramente procedimental ("procedural approach"). A raíz de ello, los autores proponen una adaptación de los sistemas de evaluación para fomentar el primer enfoque. Además, Kofinas et al. (2025) advierten sobre las limitaciones de las herramientas antiplagio para detectar de forma fiable el uso indebido de la IAG, por lo que recomiendan reformular los sistemas de evaluación en lugar de centrarse en la detección o penalización.

En el ámbito de la enseñanza y el aprendizaje de lenguas extranjeras, la literatura identifica múltiples beneficios y retos específicos. A los beneficios generales arriba mencionados se suman ventajas concretas como la oportunidad de practicar el idioma con mayor frecuencia, recibir retroalimentación personalizada, interactuar en la lengua meta y resolver dudas en tiempo real (Chen et al., 2025). Pegrum (2024) destaca enfoques innovadores como el uso de tutores socráticos impulsados por IA que estimulan la reflexión del estudiante, o la interacción con personajes virtuales en escenarios comunicativos simulados. En esta línea, Khoso et al. (2025), con una muestra de 370 estudiantes en China, concluyen que la IAG favorece la creatividad y el compromiso estudiantil. En cuanto al aprendizaje au-



torregulado - concebido como la gestión activa de los procesos cognitivos, motivacionales y conductuales orientados al logro académico (Zimmermann, 2002) -, el metaanálisis de Sardi et al. (2025) muestra que en más del 70 % de los estudios revisados la IAG tiene un impacto positivo sobre estas habilidades.

No obstante, persisten retos relevantes en el uso de la IAG para el aprendizaje de idiomas. Doğan y Talas (2025), mediante un análisis bibliométrico, señalan limitaciones como la necesidad de interacción humana, las dificultades para reproducir matices culturales y contextuales, la comprensión limitada por parte del estudiante y la dependencia de grandes volúmenes de datos. En relación con la evaluación, diversos estudios subrayan el valor del *feedback* humano —aislado o en combinación con la IAG —, pero también advierten de la escasez de investigaciones empíricas sobre su impacto real en los resultados de aprendizaje (Aliakbari et al., 2025; Li et al., 2025). En el ámbito de la evaluación formativa y sumativa (Black & Wiliam, 2009), la IAG presenta un potencial considerable, especialmente en contextos de evaluación formativa, entendida como el conjunto de prácticas destinadas a ofrecer retroalimentación con el propósito de mejorar los procesos de aprendizaje. Sin embargo, su incorporación en la evaluación sumativa —que mide el rendimiento del estudiantado en función de estándares previamente establecidos— plantea desafíos relevantes en términos de fiabilidad, validez y equidad.

En este contexto, se hace indispensable promover la alfabetización en IA (*AI Literacy*) entre docentes y estudiantes (Pegrum, 2024), con el objetivo de fomentar una conciencia crítica y un uso ético y efectivo que refuerce la autonomía del aprendizaje. No obstante, subsisten lagunas importantes en la investigación, especialmente en lo que respecta a estudios empíricos sobre el impacto de la IAG en el aprendizaje de idiomas y el diseño de intervenciones dirigidas a habilidades específicas. Tal como concluye Law (2024) en su revisión sistemática, son necesarios estudios que analicen con mayor precisión los efectos y la viabilidad pedagógica del uso de la IAG en contextos concretos de enseñanza de lenguas extranjeras.

Con el fin de contribuir a este debate, se desarrolló la intervención presentada en este artículo, cuyo objetivo es obtener datos empíricos sobre el uso de la IAG en el aprendizaje del alemán como lengua extranjera, mediante el diseño de un examen que integra el uso de dicha tecnología y el análisis de los resultados, combinado con una encuesta sobre la percepción de los participantes. La investigación plantea las siguientes preguntas:

1. ¿Qué tipo de aprendizaje puede derivarse de un examen de un idioma a libro abierto que incorpora el uso de IAG?
2. ¿Cómo percibe el alumnado esta modalidad de examen y qué necesidades formativas emergen a partir de ella?

Se espera que los resultados del estudio contribuyan al diseño de prácticas más eficaces para la integración de la IAG en los procesos de aprendizaje y evaluación en la enseñanza de lenguas extranjeras.



## 2. METODOLOGÍA

La presente experiencia de innovación docente se desarrolló en el contexto de la asignatura *Alemán B2.1* como respuesta a una situación imprevista: la imposibilidad de impartir una sesión debido a una ausencia docente que no pudo ser compensada en otro momento. Dado que dicha sesión coincidía con la fecha prevista para el examen parcial correspondiente a una lección específica, se propuso realizar dicha evaluación en un formato no presencial. Esta circunstancia operó como catalizador para explorar un enfoque alternativo de evaluación, concretamente mediante un examen a libro abierto y con la incorporación de herramientas de inteligencia artificial generativa (IAG), en el marco del aprendizaje de lenguas extranjeras.

La intervención se llevó a cabo en el contexto de una asignatura obligatoria anual de 10 créditos ECTS, correspondiente al nivel B2.1 del Marco Común Europeo de Referencia para las Lenguas (MCER), integrada en el Grado en *Global Communication*. Esta titulación se ofrece como doble grado en combinación con *Relaciones Internacionales* o *Traducción e Interpretación*. Participaron en la experiencia un total de 19 estudiantes matriculados, conformando la muestra completa de la intervención. Algunos de ellos habían cursado los niveles previos de alemán dentro del propio itinerario universitario, mientras que otros los habían completado en instituciones externas. En consecuencia, las competencias lingüísticas previas de los estudiantes presentaban un grado notable de heterogeneidad. Cabe destacar que un porcentaje elevado del grupo se encontraba en su último año académico, lo que incrementaba su preocupación por el rendimiento en la asignatura y la necesidad de superarla con éxito, especialmente dada la percepción generalizada de su alto nivel de exigencia.

La intervención se desarrolló durante el segundo semestre del curso académico, en el marco de una prueba objetiva correspondiente a la Lección 4 del programa, con un peso del 2,5 % en la calificación final de la asignatura. Dado su peso evaluativo reducido, se consideró una prueba de bajo impacto o *low stakes*. Los estudiantes fueron debidamente informados sobre el formato específico del examen y los tres objetivos principales que guiaban la actividad: (1) resolver una dificultad logística relacionada con la imposibilidad de impartir una sesión presencial; (2) comprobar la adquisición de competencias lingüísticas en alemán mediante un formato innovador de evaluación, alineado con los cambios actuales en los procesos de producción de contenido y aprendizaje mediados por inteligencia artificial generativa (IAG); y (3) fomentar una reflexión crítica sobre las capacidades, limitaciones y responsabilidad del usuario en la interacción con herramientas de IAG. En coherencia con estos objetivos, el diseño del instrumento priorizó la evaluación formativa sobre la sumativa.

El instrumento de evaluación diseñado para esta intervención (véase Anexo I) constaba de dos secciones principales, cada una compuesta por tres ejercicios. La primera sección se centraba en el uso adecuado de estructuras gramaticales y expresiones aprendidas previamente, específicamente los conectores temporales y los adjetivos terminados en *-weise*. La segunda sección estaba orientada



a la aplicación del léxico relativo al proceso de solicitud de empleo (elaboración de currículum y entrevista laboral), con énfasis en la adaptación pragmática del lenguaje a situaciones comunicativas concretas.

El diseño de las tareas integraba la producción textual en lengua alemana con la revisión crítica de los textos generados mediante inteligencia artificial generativa (IAG), exigiendo del alumnado una intervención activa sobre la forma y el contenido. Para ello, se requería que el estudiante transitara por una serie de procesos cognitivos y lingüísticos, que incluían: (1) la elaboración de *prompts* específicos que implicaban recuperar y aplicar los contenidos de la lección; (2) la reflexión sobre la pertinencia del resultado proporcionado por la IAG en relación con la tarea solicitada; (3) la verificación de la corrección formal del texto generado; y (4) la evaluación de su adecuación pragmática al contexto situacional planteado.

Se informó a los estudiantes de que, debido al carácter abierto del examen y al uso permitido de recursos como la IAG, los criterios de evaluación se ajustarían con respecto a los formatos tradicionales. Concretamente, cada error cometido supondría una penalización directa sobre la puntuación máxima de 10 puntos, lo que convertía la revisión crítica y detallada del texto en un componente esencial del rendimiento. Este enfoque buscaba incentivar una actitud crítica activa hacia la evaluación automatizada, fomentando la responsabilidad del estudiante en la validación del contenido generado.

Dado que se trataba de una primera implementación experimental de este tipo de examen, se ofreció una garantía adicional: en caso de que la media del grupo fuera significativamente inferior a la obtenida en exámenes previos realizados bajo condiciones tradicionales, se aplicaría un ajuste compensatorio para evitar posibles perjuicios. Efectivamente, esta circunstancia se dio, por lo que se ofreció una segunda oportunidad de mejora de la calificación. En esta nueva fase, el alumnado recibió una pre-corrección por parte de la docente, en la que se señalaban los errores cometidos sin proporcionar explicaciones ni soluciones. Esta estrategia generó una segunda ronda de revisión autónoma y reflexión lingüística por parte de los estudiantes.

Finalizada la intervención, se administró un cuestionario anónimo (véase Anexo II) compuesto por seis ítems. Estos abordaban: (1) el consentimiento informado para el uso de los datos recopilados; (2) la valoración de la experiencia de evaluación concreta; (3) la percepción general sobre el uso de la IAG en el aula de lenguas extranjeras; (4) sugerencias de cómo integrar la IAG en su aprendizaje, (5) la identificación de necesidades formativas en relación con el uso de estas herramientas; y (6) un campo abierto para comentarios o sugerencias adicionales. Finalmente, se procedió al análisis combinado de los errores observados en los textos evaluados y de las respuestas del cuestionario con el objetivo de responder a las dos preguntas de investigación planteadas en el estudio.

El uso de inteligencia artificial generativa (IAG) en la elaboración del presente trabajo se enmarca en el nivel 3 de la *AI Assessment Scale* propuesta por Perkins et al. (2024), el cual contempla la utilización de herramientas basadas en inteligencia artificial como apoyo en aspectos formales y de organización del texto aca-



démico. En este sentido, para la elaboración de este artículo se empleó *ChatGPT* para obtener sugerencias relativas a la estructura argumentativa, así como para la revisión lingüística y estilística del manuscrito. Asimismo, se recurrió a *Perplexity AI* como apoyo en la localización y exploración inicial de literatura relevante.

### 3. RESULTADOS

En el marco de la evaluación global de la asignatura, se administraron cuatro pruebas objetivas, cada una con un valor del 2,5 % del total, cuyos resultados se presentan en la tabla 1.

**Tabla 1. Notas medias de las pruebas objetivas**

| Prueba 1 | Prueba 2 | Prueba 3 | Prueba 4A | Prueba 4B |
|----------|----------|----------|-----------|-----------|
| 6/10     | 6.1/10   | 6.2/10   | 5.0       | 6.5       |

Pruebas 1 a 3 fueron impartidas en un formato “tradicional”: en el aula sin acceso a ninguna herramienta de apoyo. Prueba 4A fue el examen a libro abierto con uso libre de la IAG. Destacamos que la prueba 4A dio un resultado de una nota media aproximadamente un punto más bajo que la media de notas de las pruebas tradicionales.

Al comparar los resultados individuales de los alumnos en las pruebas tradicionales comparadas con la prueba con IAG (véase tabla 2), se nota una desviación estándar de 3,05, lo que implica que las diferencias individuales fueron muy dispares: algunos estudiantes bajaron su rendimiento notablemente con la IAG, mientras que otros mejoraron. Esta dispersión sugiere que el cambio de formato no tuvo un efecto uniforme, y que la manera en que cada estudiante se relacionó con la IAG fue determinante para su resultado.

**Tabla 2. Comparativa de resultados de las pruebas tradicionales con la prueba con IAG**

| Estudiantes | Pruebas tradicionales | Prueba con IAG |
|-------------|-----------------------|----------------|
| E01         | 5,5                   | 8,0            |
| E02         | 6,4                   | 4,0            |
| E03         | 3,1                   | 7,0            |
| E04         | 6,6                   | 1,0            |
| E05         | 6,0                   | 2,0            |
| E06         | 5,9                   | 5,0            |



| <b>Estudiantes</b> | <b>Pruebas tradicionales</b> | <b>Prueba con IAG</b> |
|--------------------|------------------------------|-----------------------|
| E07                | 4,6                          | 3,0                   |
| E08                | 7,2                          | 6,0                   |
| E09                | 6,0                          | 4,0                   |
| E10                | 7,9                          | 4,0                   |
| E11                | 6,1                          | 8,0                   |
| E12                | 6,5                          | 5,0                   |
| E13                | 0,0                          | 0,0                   |
| E14                | 4,8                          | 9,0                   |
| E15                | 9,4                          | 10,0                  |
| E16                | 5,8                          | 0,0                   |
| E17                | 6,8                          | 7,0                   |
| E18                | 8,8                          | 4,0                   |
| E19                | 5,6                          | 8,0                   |

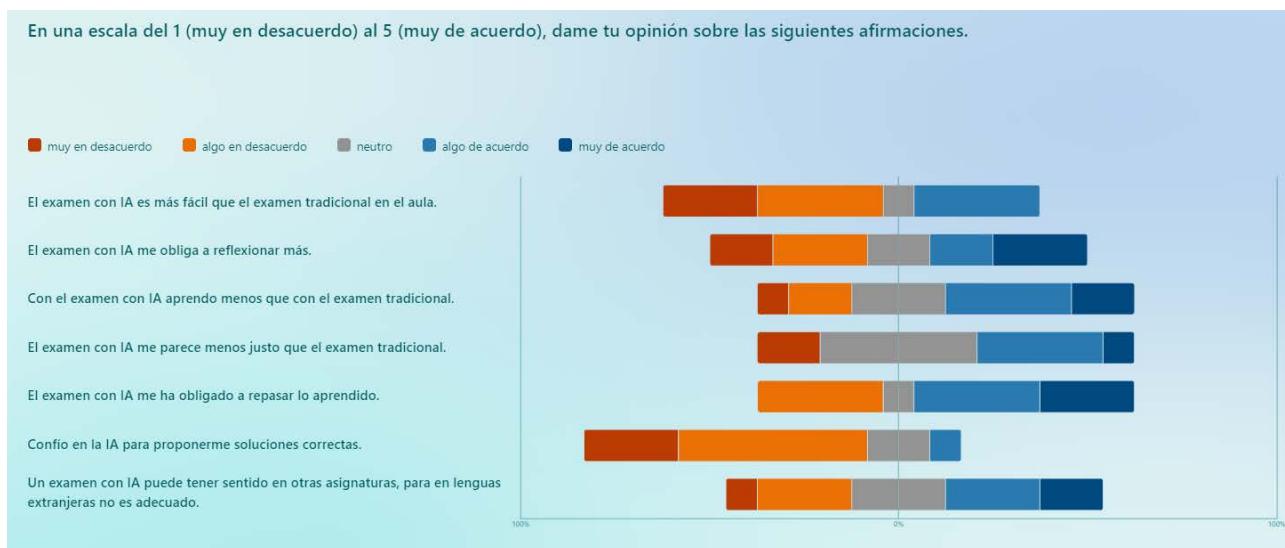
Como la media de notas fue un punto más bajo que la media de las pruebas tradicionales, se hizo una pre-corrección y se dio la oportunidad de una segunda entrega mejorada, siendo la nota media de ambas entregas la nota final de la prueba 4B (véase tabla 1).

En lo que a los errores se refiere, en nuestro análisis destaca el altísimo número de errores de lengua en los textos producidos, generalmente relacionados con la ortografía y la morfología (49 errores), seguido por un alto número de errores en las explicaciones dadas (29), la falta de seguir las instrucciones (25) y los errores relacionados con la pragmática (10).

Las percepciones del estudiantado sobre la experiencia se recogieron mediante un cuestionario de seis ítems. Doce de los 19 participantes respondieron al cuestionario y autorizaron el uso de sus datos para este estudio. Los resultados del ítem 2 (véase Figura 1) indican que el examen mediado por IAG no fue percibido como más fácil; además, el bajo nivel de confianza en las respuestas generadas por la IAG se asocia con una escasa sensación de aprendizaje y una percepción de menor justicia en la evaluación.



**Figura 1. Las opiniones de los estudiantes sobre el examen con uso de la IAG**



En el ítem 3, relacionado con el cumplimiento de las expectativas respecto al examen, la mayoría de los participantes indicó que esperaba una calificación superior a la obtenida. Varios señalaron que, en adelante, desconfiarán más de la IA. Destacan dos comentarios ilustrativos: uno de los estudiantes expresó su preferencia por un formato tradicional, argumentando que “de esa manera mi nota, tanto si es buena como mala, es mía”, lo cual refleja una sensación de pérdida de agencia y preocupación por la equidad en la evaluación. Otro participante afirmó que “con este tipo de examen los fallos pequeños cuentan más que en un examen normal, por lo que creo que el foco no es solamente el temario”, lo que pone de manifiesto su percepción de una desconexión entre el contenido evaluado y lo trabajado en clase.

Al ser consultados sobre cómo debería incorporarse la IA en la enseñanza de lenguas extranjeras, los participantes destacaron su utilidad para generar oportunidades de práctica y resolver dudas. No obstante, subrayaron la importancia insustituible de la mediación docente y la interacción humana. Como señaló un estudiante: “Los idiomas son algo muy humano que se basa en el contacto entre personas, y hay matices que una IA no puede captar.”

El ítem 5 solicitaba sugerencias sobre la formación en IA que los estudiantes considerarían útil. Entre las propuestas destacaron la elaboración de *prompts* adecuados, el uso crítico de la herramienta y su aplicación preferente en niveles avanzados de alemán, donde el estudiante puede evaluar las respuestas generadas. En el comentario libre final, solo un participante intervino, señalando que la experiencia le había resultado satisfactoria.



## 4. DISCUSIÓN

El análisis de los datos obtenidos permite formular diversas reflexiones y sugerencias en relación con las dos preguntas de investigación. La primera se centra en el aprendizaje de una lengua extranjera vinculado a un examen a libro abierto. Tanto los resultados del examen como los del cuestionario indican que este tipo de prueba favorece la reflexión y el desarrollo de habilidades receptivas, pero resulta menos eficaz para las productivas. La elevada incidencia de errores ortográficos, morfológicos y, en menor medida, pragmáticos pone de manifiesto una falta de revisión, atribuible posiblemente a un exceso de confianza en la IAG o, sencillamente, a descuido, así como a dificultades del estudiantado para identificar y corregir dichos errores. Asimismo, los problemas de comprensión de las consignas sugieren que este formato requiere una preparación docente más específica, que enfatice las estrategias necesarias para afrontarlo con éxito.

La alta desviación estándar observada en las calificaciones apunta a que el examen mediado por IAG evalúa competencias distintas a las de un examen tradicional. Podría cuestionarse su capacidad para reflejar niveles avanzados de competencia lingüística, ya que parece medir habilidades como la atención al detalle, la conciencia crítica y la capacidad de adaptación a nuevos formatos, más que el dominio efectivo del alemán.

La tipología de errores observada en el examen y las conclusiones derivadas de ellos se reflejan en las respuestas al cuestionario, donde la mayoría de los participantes considera que este formato les proporciona un aprendizaje menor en comparación con un examen tradicional. Asimismo, señalan que muchos de sus errores se vinculan con aspectos que no fueron centrales en los contenidos abordados durante la lección objetivo de la prueba. Esta observación remite directamente a una cuestión ética fundamental: la justicia en la evaluación. En un estudio comparativo, Ye et al. (2024) reportan resultados similares, señalando que tanto el profesorado como el estudiantado manifiestan preocupación por la equidad ("fairness") de los exámenes a libro abierto. La falta de alineación entre los contenidos trabajados en clase y los aspectos evaluados genera una percepción de inequidad. Esta discrepancia evidencia una limitación en el diseño de la prueba. De cara a futuras implementaciones, sería recomendable revisar y ajustar los criterios de evaluación, de modo que los errores ajenos al temario específico tengan un peso menor en la calificación final.

La posibilidad de una segunda revisión y entrega del examen resultó fundamental tanto para que el estudiantado percibiera la evaluación como más justa, como para favorecer un aprendizaje más profundo a partir de sus errores. Esta experiencia refuerza, una vez más, la importancia de la retroalimentación individualizada como elemento clave del proceso formativo, independientemente de que se trate de formatos de evaluación tradicionales o innovadores.

La segunda pregunta de investigación se centró en la percepción del estudiantado sobre el uso de la inteligencia artificial generativa (IAG) en la docencia, y en particular, en los procesos de evaluación, así como en sus necesidades formati-



vas. Las respuestas al cuestionario evidencian un cambio en la percepción de la dificultad relativa entre un examen tradicional y uno mediado por IAG. Los participantes muestran una mayor conciencia de que la IAG no es infalible y que sus resultados requieren supervisión humana para ser fiables. Si bien reconocen su utilidad para generar ejercicios, resolver dudas o facilitar la práctica autónoma, manifiestan una clara desconfianza hacia su uso como herramienta evaluativa. Asimismo, destacan la necesidad de recibir formación específica antes de integrar la IAG en contextos académicos. Estos hallazgos coinciden con lo señalado en la literatura reciente (cf. Mirea et al., 2025; Niño-Carrasco et al., 2025).

## 5. CONCLUSIONES

Tras la implementación de un examen a libro abierto en el contexto del aprendizaje de una lengua extranjera, se evidencia la necesidad de una formación previa tanto para el profesorado como para el estudiantado, con el fin de garantizar que este tipo de instrumento evalúe adecuadamente los resultados de aprendizaje propuestos. La retroalimentación individualizada emerge como un componente esencial para consolidar dicho aprendizaje. En consecuencia, se concluye que el uso de la inteligencia artificial generativa (IAG) resulta más adecuado en contextos de evaluación formativa, donde puede apoyar el desarrollo de competencias de manera progresiva. En cambio, en el ámbito de la evaluación sumativa, esta intervención sugiere que un examen con uso de la IAG no permite medir con fiabilidad las competencias lingüísticas específicas requeridas. Por ello, se recomienda complementar su uso con instrumentos presenciales en los que el acceso a herramientas de IA sea controlado o restringido, asegurando así la validez del proceso evaluativo.

El presente estudio constituye una intervención puntual, desarrollada en el marco de una iniciativa de innovación docente. No fue concebido como un diseño de investigación sistemático; en consecuencia, el número limitado de participantes y la naturaleza exploratoria de los datos no permiten generalizar los resultados a otros contextos educativos. No obstante, la experiencia ha generado una serie de recomendaciones de buenas prácticas para futuras implementaciones:

1. Incorporar un diseño mixto de examen que combine secciones a libro abierto con otras en las que se controle la autoría del estudiante.
2. Diferenciar claramente los objetivos de aprendizaje lingüístico de aquellos vinculados al uso pedagógico de la IAG.
3. Ampliar y sistematizar la formación previa del estudiantado en el uso crítico y ético de la IAG.
4. Integrar un ciclo de retroalimentación formativa previo a la evaluación final, con el fin de reforzar el aprendizaje y mejorar la percepción de justicia evaluativa.



Esta experiencia de innovación docente pone de manifiesto los retos que implica el diseño de exámenes mediados por inteligencia artificial generativa (IAG) en contextos de aprendizaje de lenguas extranjeras. Los resultados evidencian que, si bien la IAG puede ofrecer oportunidades valiosas para la práctica y el desarrollo de habilidades receptivas, su uso como instrumento de evaluación plantea importantes limitaciones en términos de fiabilidad, justicia y alineación con los objetivos de aprendizaje. Tanto el profesorado como el estudiantado requieren formación específica para integrar esta tecnología de manera crítica y pedagógicamente coherente (cf. Krause et al., 2025). En suma, aún queda un importante camino por recorrer para que la IAG pueda incorporarse de forma efectiva y ética en los procesos evaluativos.

## AGRADECIMIENTOS

Quiero expresar mi agradecimiento al **Servicio de Apoyo a la Innovación Docente de la Universidad Pontificia Comillas** por su constante labor de acompañamiento y por haber facilitado la oportunidad de desarrollar y difundir este trabajo, el cual fue distinguido con el segundo premio en el **Reto Docente Innovador 2024-25**.

## REFERENCIAS

- Aliakbari, M., Barzan, P., & Sayyadi, M. (2025). Automated feedback vs. human feedback: A study on AI-driven language assessment. *AI and Tech in Behavioral and Social Sciences*, 3(2), 113 - 126. <https://journals.kmanpub.com/index.php/aitechbesosci/article/view/3813>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment. Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Chen, Z., Wei, W., & Zou, D. (2025). Generative AI technology and language learning: Global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, 1–14. <https://doi.org/10.1080/10494820.2025.2511248>
- Doğan, Y., & Talan, T. (2025). Review Article: Artificial intelligence in foreign language learning. *Journal of Pedagogical Research*, 9(2), 206-230. <https://doi.org/10.33902/JPR.202427734>
- Fundación CYD. (2025, 7 de mayo). *La Fundación CYD analiza el uso de la IA en el entorno universitario*. <https://www.fundacioncyd.org/la-fundacion-cyd-analiza-el-uso-de-la-ia-en-el-entorno-universitario/>
- Higher Education Policy Institute. (2025). *Student generative AI survey 2025*. <https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/>
- Jisc. (2025). *Student perceptions of AI 2025*. <https://www.jisc.ac.uk/reports/student-perceptions-of-ai-2025>



- Khoso, A.K., Wang, H.G., & Darazi, M.A. (2025). Empowering creativity and engagement: The impact of generative artificial intelligence usage on Chinese EFL students' language learning experience. *Computers in Human Behavior Reports*, 18 (100627), 1-14. <https://doi.org/10.1016/j.chbr.2025.100627>
- Kofinas, A. K., Tsay, C.H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments in higher education. *British Journal of Educational Technology*, 00, 1-28. <https://doi.org/10.1111/bjet.13585>
- Krause, S., Panchal, B.H., & Ubhe, N. (2025). Evolution of learning: assessing the transformative impact of generative AI on higher education. *Frontiers of Digital Education*, 2(21), 1-15. <https://doi.org/10.1007/s44366-025-0058-7>
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6(100174), 1-13. <https://doi.org/10.1016/j.caeo.2024.100174>
- McDonald, N., Johri, A., Ali, A., & Hingle Collier, A. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3(100121), 1-11. <https://doi.org/10.1016/j.chbah.2025.100121>
- Mirea, C.-M., Bologa, R., Toma, A., Clim, A., Plăcintă, D.-D., & Bobocea, A. (2025). Transformar el aprendizaje con IA generativa: de las percepciones de los estudiantes al diseño de una solución educativa. *Applied Sciences*, 15(10), 5785. <https://doi.org/10.3390/app15105785>
- Niño-Carrasco, S. A., Castellanos-Ramírez, J. C., Perezchica Vega, J. E., & Sepúlveda Rodríguez, J. A. (2025). Percepciones de estudiantes universitarios sobre los usos de inteligencia artificial en educación. *Revista Fuentes*, 27(1), 94–106. <https://doi.org/10.12795/revistafuentes.2025.26356>
- Pallant, J. L., Blijlevens, J., Campbell, A., & Jopp, R. (2025). Mastering knowledge: The impact of generative AI on student learning outcomes. *Studies in Higher Education*, 1-22. <https://doi.org/10.1080/03075079.2025.2487570>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- Sardi, J., Candra, D. O., Yuliana, D. F., Yanto, H. D. T. P., & Eliza, F. (2024). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy (iJEP)*, 15(1), 94–108. <https://doi.org/10.3991/ijep.v15i1.53379>
- Ye, X., Yang, Y., Qie, Y. (2024). The impact of 24-h take-home exam on language learning and teaching on the China campus of a British university. *Language Testing in Asia*, 14(23), 1-27. <https://doi.org/10.1186/s40468-024-00295-z>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)



## ANEXO I

TEST B 2.1

### Sicher B2 Lektion 4

*Take Home Prüfung*

**[Bearbeitungszeit: frei]**

Name(n): .....

Bewertung:

Jeder Fehler und jede Auslassung reduziert die Note um 1 Punkt.

Diese Inhalte können in der mündlichen Prüfung im April abgefragt werden.

#### 1. Stand 27.3.2025 **Temporale Konnektoren: Satzumformung**

Verbinden Sie die Sätze mit temporalen Konnektoren. Machen Sie alle notwendigen Änderungen, um korrekte Sätze zu bilden. Erklären Sie Ihre Entscheidungen.

| Aussgangssätze   | Satzkombination mit temporalem Konnektor | Erklärung (deutsch oder spanisch) |
|--|--|-----------------------------------|
| 1. Ich habe gefrühstückt. Dann bin ich zur Arbeit gegangen. → Verwenden Sie: <b>nachdem</b>            |  |                                   |
| 2. Ich lerne Deutsch. Ich höre Musik. → Verwenden Sie: <b>während</b> (als Konjunktion)                |  |                                   |
| 3. Ich möchte meine Arbeitsstelle wechseln. Ich finde eine neue Stelle. → Verwenden Sie: <b>sobald</b> |  |                                   |
| 4. Er wurde krank. Er arbeitete Tag und Nacht. → Verwenden Sie: <b>bevor</b>                           |  |                                   |
| 5. Ich war ein Kind. Ich fuhr jeden Sommer nach Spanien in Urlaub. → Verwenden Sie: <b>als/wenn</b>    |  |                                   |

#### 2. Microschreiben

Schreiben Sie 5 Sätze über Ihren Tagesablauf auf der Arbeit oder im Praktikum. Benutzen Sie mindestens 4 temporale Konnektoren aus Lektion 4, von denen 2 verbal und 2 nominal sind. Erklären Sie kurz, wie die Struktur funktioniert.

##### **Beispiel:**

Bevor ich den Computer einschalte, trinke ich schnell einen Kaffee. [Ich benutze eine verbale Konstruktion: Konjunktion „bevor“ + Nebensatz, Hauptsatz. Bei „bevor“ sind die Verbzeiten in beiden Teilen des Satzes gleich: Präsens. Der NS steht auf Position 1, also steht im HS das Verb direkt nach dem Komma, auf Position 2.]




### 3. Adverbien mit “-weise” (KB, S. 60 & AB, S. 70)

**Korrigieren Sie die Fehler und erklären Sie sie.**

| Satz mit Fehler(n)  | Erklärung (DE oder ES) | Korrektur |
|---|------------------------|-----------|
| 1. Dieses Jahr macht unsere Lehrerin proberweise eine <i>Take Home</i> Klausur mit uns.                   |                        |           |
| 2. Viele Studierende haben verständlichweise keine Zeit, zwei Bachelorarbeiten gleichzeitig zu schreiben. |                        |           |
| 3. Markus hat eine Allergie gegen Pollen und musste heute erfreulichweise zum Notarzt.                    |                        |           |

### 4. Lebenslauf

Nennen Sie 5 typische Teile eines Lebenslaufs auf Deutsch.

### 5. Bewerbungsgespräch (1)

Schreiben Sie die Antworten auf diese Fragen in Bezug auf die Stellen, die Sie für das Examen ausgewählt haben und passend zu Ihrem Lebenslauf.

**Interviewer:** Warum möchten Sie bei uns arbeiten?

**Sie:** \_\_\_\_\_

**Interviewer:** Was sind Ihre Stärken und Schwächen?

**Sie:** \_\_\_\_\_

**Interviewer:** Wo sehen Sie sich in fünf Jahren?

**Sie:** \_\_\_\_\_

### 6. Bewerbungsgespräch (2)

Was ist falsch an der untenstehenden Antwort? Schlagen Sie eine bessere Antwort vor, in Bezug auf die Stelle, die Sie für das Examen ausgewählt haben. Benutzen Sie mindestens 2 Ausdrücke aus dem KB, S. 59.

**Frage:** Warum haben Sie sich bei uns beworben?

**Antwort:** Keine Ahnung, mein Freund hat gesagt, ich soll.



## ANEXO II

Enlace a la encuesta: <https://forms.office.com/e/zjfwYHhK8g?origin=lprLink>

### Take Home Prüfung: Feedback

Como es la primera vez que he hecho un examen de llevar a casa y con uso de la IA y cualquier otra fuente de información, me gustaría que me comentarais vuestras impresiones. Por favor, dadme vuestra opinión con toda honestidad. Es un cuestionario anónimo y sólo me sirve para valorar el resultado del examen y compartirlo con otros compañeros que también han empezado a experimentar con la IA.

1. Doy mi permiso para que los resultados de este cuestionario se compartan con otros profesores, de forma anonimizada.

- Si  
 No

2. En una escala del 1 (muy en desacuerdo) al 5 (muy de acuerdo), dame tu opinión sobre las siguientes afirmaciones.

|  | muy en desacuerdo     | algo en desacuerdo    | neutro                | algo de acuerdo       | muy de acuerdo        |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| El examen con IA es más fácil que el examen tradicional en el aula.                                    | <input type="radio"/> |
| El examen con IA me obliga a reflexionar más.  | <input type="radio"/> |
| Con el examen con IA aprendo menos que con el examen tradicional.                                      | <input type="radio"/> |
| El examen con IA me parece menos justo que el examen tradicional.                                      | <input type="radio"/> |
| El examen con IA me ha obligado a repasar lo aprendido.  | <input type="radio"/> |
| Confío en la IA para proponerme soluciones correctas.  | <input type="radio"/> |
| Un examen con IA puede tener sentido en otras asignaturas, para en lenguas extranjeras no es adecuado. | <input type="radio"/> |



3. Después del examen, cuéntame tu experiencia. ¿Cuáles habían sido tus expectativas? ¿Se cumplieron? ¿Qué sensación tienes después del examen?

Escriba su respuesta

4. ¿De qué forma crees que la IA se podría integrar de forma constructiva en la enseñanza de idiomas?

Escriba su respuesta

5. ¿Qué formación necesitarías para poder utilizar la IA de forma eficaz en el aprendizaje de lenguas extranjeras?

Escriba su respuesta

6. Gracias por tu colaboración. Si tienes cualquier otro comentario que quieras hacer, puedes escribirlo aquí.

Escriba su respuesta