

## *Del lápiz al píxel: Gradescope como herramienta de evaluación de exámenes en ciencias exactas, prueba piloto en matemáticas*

### **From pencil to pixel: Gradescope as an evaluation tool for exams in exact sciences, a pilot test in mathematics**

Glauco Alfredo López Díaz<sup>\*1</sup> , Carlos Iván Zelaya Valeriano<sup>1</sup> 

<sup>1</sup>Departamento de Educación General, Universidad Zamorano, San Antonio de Oriente, Honduras

\*Autor correspondiente: [glopez@zamorano.edu](mailto:glopez@zamorano.edu)



Este trabajo está bajo una licencia internacional Creative Commons Attribution 4.0 BY, NC.

Recepción: 16/12/2025 Aceptación: 03/02/2026 Publicación: 02/05/2026

Cita: López Díaz, G. A., Zelaya Valeriano, C. I. (2026). Del lápiz al píxel: Gradescope como herramienta de evaluación de exámenes en ciencias exactas, prueba piloto en matemáticas. *Lumen, Revista de Innovación Educativa* 1(1), 1-6.

#### **Palabras clave:**

Evaluación de la educación,  
Evaluación del estudiante,  
Inteligencia artificial, Método de evaluación,  
Tecnología educacional

#### **RESUMEN**

Contribuir con la mejora en la evaluación de trabajos de los estudiantes, especialmente cuando son procedimientos matemáticos siempre es un reto. La plataforma en línea Gradescope optimiza la calificación de evaluaciones escritas y de respuesta libre utilizando una rúbrica dinámica y ajustable durante el proceso de evaluación. La herramienta está diseñada para gestionar exámenes, tareas y pruebas cortas, facilitando una organización eficiente, calificación consistente y retroalimentación estructurada. Se integra con sistemas de gestión del aprendizaje como Blackboard Ultra, simplificando la sincronización de listas de alumnos y la transferencia de calificaciones. Para estudiar la efectividad se implementó un plan piloto de Gradescope en el Departamento de Educación General de la Universidad Zamorano. La prueba se centró en la evaluación automática del cuarto examen parcial de Álgebra Lineal. El proceso incluyó la configuración del curso y la carga de exámenes escaneados para su evaluación con una rúbrica digital. Los datos obtenidos de la experiencia demuestran una reducción del 50% en el tiempo de calificación comparado con métodos manuales (revisión con lápiz). Los estudiantes accedieron a sus notas y retroalimentación en línea de forma instantánea. La automatización (revisión a píxel) eliminó errores humanos en la totalización e interpretación de resultados, despersonalizando el proceso y evitando discusiones posteriores. Aun cuando, la configuración inicial requiere invertir tiempo, el beneficio en eficiencia y precisión durante la corrección subsiguiente es significativo. Gradescope agiliza la gestión de evaluaciones a gran escala, proporcionando una solución robusta para la docencia de ciencias.

#### **Keywords:**

Artificial intelligence,  
Educational evaluation,  
Educational technology,  
Evaluation method,  
Student assessment

#### **ABSTRACT**

Contributing to the improvement of student work assessment, especially when it involves mathematical procedures, is always a challenge. The Gradescope online platform optimizes the grading of written and free-response assessments through a dynamic rubric that can be adjusted during the assessment process. The tool is designed to manage exams, assignments, and short tests, facilitating efficient organization, consistent grading, and structured feedback. It integrates with learning management systems such as Blackboard Ultra, simplifying the synchronization of student lists and the transfer of grades. To study its effectiveness, a Gradescope pilot program was implemented in the General Education Department at Zamorano University. The test focused on the automatic evaluation of the fourth midterm exam in Linear Algebra. The process included setting up the course and uploading scanned exams for evaluation with a digital rubric. The data obtained from the experience shows a 50% reduction in grading time compared to manual methods (pencil review). Students accessed their grades and feedback online instantly. Automation (pixel review) eliminated human error in the tabulation and interpretation of results, depersonalizing the process and avoiding subsequent discussions. Although the initial setup requires an investment of time, the benefits in terms of efficiency and accuracy during subsequent grading are significant. Gradescope streamlines the management of large-scale assessments, providing a robust solution for science teaching.

## INTRODUCCIÓN

En la era digital, la educación superior experimenta una constante evolución impulsada por la necesidad de integrar tecnologías que mejoren la enseñanza y el aprendizaje. En este contexto, la evaluación es fundamental, ya que no solo mide el conocimiento adquirido, sino que también proporciona retroalimentación crucial para el desarrollo académico de los estudiantes (Rodríguez-Vílchez, 2019; Okafor, 2025). Las metodologías tradicionales de calificación manual de exámenes escritos, especialmente en disciplinas cuantitativas como las matemáticas, consumen tiempo docente considerable y son susceptibles a errores humanos, afectando la consistencia y la inmediatez de la retroalimentación (Bloxham & Boyd, 2007).

La transición hacia plataformas digitales de gestión de evaluaciones busca mitigar estas ineficiencias. En el contexto centroamericano Camacho et al. (2020) y específicamente en Honduras, la adopción de estas tecnologías en cursos de ciencias es un área de investigación emergente donde la eficiencia operativa puede impactar positivamente la carga académica del profesorado y la experiencia del alumnado (López-Silva, 2022). En este sentido, la dificultad en la evaluación en el área de las ciencias es considerable, por lo que la herramienta Gradescope optimiza estas debilidades al ofrecer características innovadoras para la gestión de exámenes y tareas, y para la evaluación, en diferentes áreas, entre estas las matemáticas.

Esta herramienta es capaz de manejar tanto ítems de opción múltiple, así como problemas de desarrollo de respuesta libre con rúbrica dinámica, permitiendo a los educadores implementar un enfoque más flexible y eficiente (Henderson et al., 2019). La plataforma no solo agiliza el proceso de corrección mediante una rúbrica dinámica, asimismo, mejora la retroalimentación que reciben los estudiantes, permitiendo un aprendizaje más profundo y significativo (Rinaldi & Sotelo, 2024). A pesar de su uso extendido en otras latitudes, la evidencia empírica sobre su implementación y efectividad en instituciones hondureñas, como la Universidad Zamorano, permanece limitada.

En este estudio se exponen los resultados de la implementación del piloto de Gradescope en el Departamento de Educación General, centrándose en el curso de Álgebra Lineal, presentando un balance de los hallazgos principales. Esta tecnología toma énfasis en un juicio comparativo, que el docente programa al inicio de la evaluación, el cual, es una alternativa fiable y válida a la calificación tradicional para evaluar las matemáticas, en particular para tareas abiertas de resolución de problemas. Este tipo de juicios produce resultados comparables a los métodos tradicionales, que, al liberar a los docentes de los esquemas de calificación prescriptivos, permite una mejor evaluación de las variadas respuestas de los estudiantes (Bisson et al., 2016).

## PRESENTACIÓN DE LA EXPERIENCIA

El objetivo general de este estudio fue evaluar la eficiencia de la plataforma de inteligencia artificial (IA) Gradescope en las evaluaciones escritas de los cursos de ciencias en el Departamento de Educación General de la Universidad Zamorano, midiendo específicamente la reducción del tiempo de calificación y la precisión de los resultados obtenidos del cuarto examen parcial del curso de Álgebra Lineal del primer periodo del 2025.

La metodología se centró en la implementación de Gradescope, una herramienta de calificación en línea para evaluaciones en papel y lápiz, escaneadas, recomendado por estudios preliminares (Schinske & Tanner, 2012). El método de estudio es cualitativo, de tipo exploratorio. El tamaño de la muestra aleatoria fue de 40 estudiantes de primer año de ingeniería.

El flujo de trabajo adoptado incluyó una serie de pasos secuenciales: creación de una cuenta y curso de instructor, carga de la lista de estudiantes, creación de la actividad, elaboración y carga de la rúbrica de evaluación y los exámenes escaneados. El proceso clave implicó agrupar respuestas similares por pregunta, ponderarlas según la rúbrica y evaluarlas automáticamente. Finalmente, se agregaron comentarios estandarizados para la retroalimentación individual antes de publicar los resultados.

La implementación de Gradescope resultó en una revisión automatizada de los exámenes, generados por el software Píxel, significativamente más veloz, disminuyendo el tiempo de evaluación a la mitad, lo cual incidió en el ahorro de una hora frente a las dos que normalmente tarda el proceso de calificación cuando se realizan totalmente de forma manual o con el lápiz. Esta herramienta, de utilidad para los docentes, también aporta beneficios para los estudiantes, quienes tienen acceso para revisar sus ejercicios y notas en cualquier momento dentro de la plataforma, lo cual refuerza el aprendizaje autónomo.

Como resultado de este flujo de trabajo, los estudiantes recibieron retroalimentación mediante comentarios estandarizados, automáticos y oportunos. Al poder acceder a sus ejercicios y notas de forma instantánea en la plataforma, se refuerza el aprendizaje autónomo y se reduce significativamente la necesidad de reclamos por interpretaciones subjetivas de la evaluación. La carga previa de la rúbrica minimizó las discusiones sobre la interpretación de criterios y la totalización automática de calificaciones eliminó el error humano en la suma de puntos. El grupo de respuesta correcta obtuvo una retroalimentación "con buena notación" como lo muestra la Figura 1, o "buena aritmética con mala notación" como lo muestra la Figura 2.

$$a) \quad A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (\text{valor } 3 \%)$$

$$|A| = \begin{vmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{vmatrix} = 3 \times 1 \times 0 + 1 \times 0 \times 1 + 1 \times 2 \times 1 - 1 \times 1 \times 1 \\ - 1 \times 0 \times 3 - 0 \times 2 \times 1 \\ = 0 + 0 + 2 - 1 - 0 - 0 \\ = 1$$

$$|A| = 1$$

Figura 1. Grupo de respuesta correcta "con buena notación" del ejercicio 1.

$$a) \quad A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (\text{valor } 3 \%)$$

$$\begin{matrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{matrix}$$

$$0 + 0 + 2 - 1 - 0 - 0 = 1 = A$$

Figura 2. Grupo de respuesta correcta "buena aritmética con mala notación" del ejercicio 1.

Comparando el método tradicional de revisión de exámenes, la experiencia con el instrumento redujo el tiempo de corrección a la mitad, eliminó errores en totalización de notas y proporcionó retroalimentación inmediata y accesible.

La estandarización de rúbricas minimizó disputas por interpretaciones subjetivas, de tal manera que los reclamos disminuyeron en un 20%, siendo la experiencia que en grupos de 40 estudiantes alrededor de ocho realizaban reclamos sobre sus calificaciones, con la implementación de este método solamente dos de 40 alumnos solicitaron revisión, lo que equivale al 5%.

En el proceso formativo los estudiantes obtuvieron una retroalimentación más detallada y al instante, al colocar el comentario de los errores obtenidos y proporcionarles la respuesta correcta, pudieron observar sus resultados cuantas veces lo necesitaran o quisieran, utilizándolo como una autocorrección que facilita el estudio y fortalece sus conocimientos. Se destaca que 32 de 40 estudiantes (80%) del grupo de muestreo consideraron que con la implementación de la herramienta lograron ser evaluados de una forma más objetiva y eficaz.

## DISCUSIÓN

La implementación de herramientas de evaluación automatizada responde a la necesidad de una educación más dinámica y centrada en el estudiante. Mientras Sangwin (2013) destaca el papel de la evaluación como impulsor del aprendizaje, Rama (2021) subraya que la educación híbrida actual exige una automatización inteligente de los procesos. Este enfoque se alinea con los componentes de la Educación 4.0 en la enseñanza de la ingeniería, donde la integración tecnológica es fundamental para la formación técnica contemporánea (Miranda et al., 2021).

En este sentido, plataformas como Gradescope permiten superar las limitaciones de los métodos estáticos como los esquemas de calificación rígidos, procesos lineales y retroalimentación fija; cumpliendo con las demandas de calidad y desarrollo profesional que Zabalza (2017) identifica como críticas para el profesorado universitario moderno. Los resultados de este estudio corroboran que las metodologías tradicionales de calificación manual consumen grandes cantidades de tiempo a los docentes y son susceptibles a errores, afectando la consistencia del proceso (Bloxham & Boyd, 2007).

La optimización operativa registrada de un ahorro del 50% en el tiempo de corrección, se sitúa dentro de los parámetros de Singh et al. (2017) y representa un avance significativo para la educación superior en Honduras. Como señala López-Silva (2022), la adopción de estas tecnologías en cursos de ciencias es un área emergente que impacta positivamente la carga académica del profesorado y la experiencia del alumnado en la región.

Finalmente, el sistema no solo agiliza el proceso, asimismo, refuerza el aprendizaje autónomo mediante una retroalimentación automática y oportuna (Dawson, 2017). El uso de rúbricas transparentes, según Panadero y Jonsson (2013), reduce la ansiedad del estudiante y proporciona una estructura clara que minimiza los reclamos posteriores, un efecto observado en esta prueba piloto donde los reclamos disminuyeron drásticamente y la accesibilidad a la retroalimentación se convierte en fuente de autocorrección y aprendizaje para los estudiantes.

## CONCLUSIONES

Se concluye que, de acuerdo con el objetivo general de evaluar la eficiencia de esta herramienta de IA, si bien existe una inversión inicial de tiempo en la configuración del curso y la plataforma, se logró implementar Gradescope con una significativa eficiencia operativa en el curso de Álgebra Lineal. La automatización del proceso facilitó que el tiempo de evaluación se redujera significativamente en un 50%, solamente en el proceso de la revisión, lo que equivale a una mejora cuantitativa de un ahorro de una hora por cada bloque de evaluación de dos horas.

En cuanto la implementación se determinó que la plataforma es altamente precisa y efectiva gracias a la funcionalidad de agrupar respuestas por similitud, lo que permite: despensar el proceso de evaluación, garantizando una mayor objetividad al centrarse en el procedimiento técnico del estudiante; estandarizar la retroalimentación, proporcionando comentarios consistentes y oportunos que disminuyeron los reclamos de un 20% a un 5%; y fortalecer la percepción estudiantil, donde el 80% de los alumnos manifestó que el proceso fue más eficaz y transparente.

La prueba piloto demostró que Gradescope constituye una solución robusta, escalable y eficiente para la gestión de evaluaciones a gran escala en el ámbito de las ciencias exactas, logrando un equilibrio entre la precisión académica en la evaluación y la optimización del tiempo invertido por el profesorado, razón por la cual, se recomienda expandir su uso a otras asignaturas con alto contenido de desarrollo práctico.

#### ***Contribución de los autores***

GALD y CIZV contribuyeron de igual forma en el planteamiento, desarrollo y documentación del caso de estudio. Ambos autores están de acuerdo con la versión final del documento.

#### ***Financiamiento***

Ninguno.

#### ***Conflictos de interés***

Ninguno.

#### ***Aprobación ética***

Exento.

#### ***Uso de IA***

Se utilizó la herramienta de IA Google (2025), Gemini, para reducir el resumen y su posterior traducción al inglés.

## **REFERENCIAS**

- Bisson, M. J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring Conceptual Understanding Using Comparative Judgement. *International Journal for Technology in Mathematics Education*, 23(4), 157–163. <https://doi.org/10.1007/s40753-016-0024-3>
- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. Open University Press.
- Camacho, R., Rivas, C., Gaspar, M., & Quiñonez, C. (2020). Innovación y tecnología educativa en el contexto actual latinoamericano. *Revista de Ciencias Sociales (Ve)*, XXVI (Especial 2), 460-472. <https://www.redalyc.org/articulo.oa?id=28064146030>
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- Henderson, M., Ryan, T., & Phillips, M. (2019). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education*, 44(8), 1237–1252. <https://doi.org/10.1080/02602938.2019.1599815>
- Miranda, J., Navarrete, C., Noguez, J., Molina-Espinosa, J. M., Ramírez-Montoya, M. S., Navarro-Tuch, S. A., Bustamante-Bello, M. R., Rosas-Fernández, J. B., & Molina, A. (2021). The core components of education 4.0 in higher education: Three case studies in engineering education. *Computers & Electrical Engineering*, 93, Article 107278. <https://doi.org/10.1016/j.compeleceng.2021.107278>
- Okafor, J. (2025). The role of digital tools in assessment and their impact on educational practices. *International Journal of Applied Research and Sustainable Sciences*, 3(1), 1–18. <https://doi.org/10.64420/ijitl.v2i1.202>

- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Rama, C. (2021). *La nueva educación híbrida*. Unión de Universidades de América Latina y el Caribe. <http://dspaceudual.org/handle/Rep-UDUAL/202>
- Rinaldi, S., & Sotelo, D. (2024). *Algunas razones (poderosas) para incorporar las inteligencias artificiales a la enseñanza universitaria* (Documento de trabajo N.º 881). Universidad del CEMA. <https://cl2.ucema.edu.ar/sites/default/files/2024-08/dt881.pdf>
- Rodríguez-Vílchez, J. A. (2019). *Evaluación formativa y retroalimentación en la universidad*. Narcea Ediciones.
- Sangwin, C. J. (2013). *Computer aided assessment of mathematics*. Oxford University Press.
- Schinske, J. N., & Tanner, K. D. (2012). Teaching minds to think: Gradescope and the future of science education. *CBE—Life Sciences Education*, 11(4), 322–324. <https://doi.org/10.1187/cbe.12-10-0166>
- Singh, A., Karayev, S., Guttenberg, N., & Abbeel, P. (2017). Gradescope: A fast, flexible, and fair system for scalable assessment of handwritten work. *Proceedings of the Fourth ACM Conference on Learning @ Scale*, 81–90. <https://doi.org/10.1145/3051457.3051466>
- Zabalza, M. A. (2017). *Competencias docentes del profesorado universitario: Calidad y desarrollo profesional*. Narcea Ediciones.