

INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN UNIVERSITARIA: EVIDENCIAS DEL USO DE HONOR EDUCATION

Ana Hidalgo-Cabrillana

Universidad Autónoma de Madrid

Resumen

Este trabajo analiza la relación entre el uso de inteligencia artificial (IA) para fomentar la lectura del libro de CORE Econ y el aprendizaje de los estudiantes. Se recopilieron notas de exámenes en dos grupos que recibieron el mismo contenido: misma asignatura, profesor y exámenes. Sin embargo, solo uno utilizó IA a través de la aplicación Honor Education (HE). Al comparar los resultados, se observa que en uno de los exámenes, centrado en evaluar la capacidad de razonamiento, el grupo que usó HE obtuvo una media significativamente superior. No ocurre lo mismo en los tres exámenes que evaluaban conocimiento y aplicación de problemas rutinarios.

Palabras clave: libro CORE Econ, aprendizaje de estudiantes, conocimiento, problemas rutinarios, razonamiento, Honor Education.

Abstract

This paper analyzes the relationship between the use of artificial intelligence (AI) to encourage reading of the CORE Econ textbook and student learning outcomes. Exam scores were collected from two groups that received the same content: same course, instructor, and exams. However, only one group used AI through the Honor Education (HE) application form. Comparing the results, we find that in one of the exams—focused on assessing reasoning skills—the group using HE achieved a significantly higher average score. This is not the case in the three exams that evaluated knowledge and the application of routine problems.

Keywords: CORE Econbook, student learning, knowledge, routine problems, reasoning.

JEL classification: A23, I23.

I. INTRODUCCIÓN

A medida que la inteligencia artificial (IA) se integra cada vez más en nuestras vidas, ¿se benefician realmente los alumnos con un mejor aprendizaje de su integración en las aulas?, ¿qué recursos son los más efectivos potenciando el aprendizaje? Y una vez identificados esos recursos, ¿de qué forma sucede? Entender estas preguntas son cruciales, pues sabemos que, por un lado, rechazar o detener la implementación de la IA en las aulas es una estrategia sin sentido, y por otro, mayores niveles de capital humano tienen beneficios no solamente individuales (por ejemplo, mejorando condiciones en el mercado de trabajo [Murname *et al.*, 1995; Rivkin *et al.*, 2005]), sino también agregados para la economía (por ejemplo, incentivando el crecimiento económico

[Hanushek y Kimko, 2000; Hanushek y Woessman, 2012]).

La aparición de la inteligencia artificial en la universidad y en niveles educativos inferiores está revolucionando tanto la forma en la que los docentes enseñan como la forma en la que los estudiantes trabajan para mejorar su aprendizaje. Para los docentes, la inteligencia artificial puede ayudar, por ejemplo, en la creación de contenido, en automatizar la evaluación a los estudiantes, o en la adaptación del contenido a las necesidades específicas de los estudiantes. Para los estudiantes, la inteligencia artificial, al tener capacidad de comprender el significado y la intención de las preguntas, ofrece una herramienta de apoyo en su proceso de aprendizaje al facilitar respuestas a los estudiantes personalizadas y contextualizadas. Además, esas preguntas

pueden ser aclaradas siguiendo las instrucciones del estudiante. Sin embargo, la utilización de la inteligencia artificial puede afectar la formación de los estudiantes al eludir el proceso de aprendizaje, fomentar la dependencia y el estudio superficial, así como la correcta evaluación. Ejemplo de actividades en el campo de la economía en las que esto puede ocurrir puede ser el análisis, la interpretación y la generación de texto, como revisión de la literatura, o casos de estudio. O aquellas actividades como el cálculo de indicadores socioeconómicos, programación o identificación y verificación de hipótesis. El uso de estas herramientas sin un pensamiento crítico puede propiciar la utilización de información inexacta, sesgada o errónea. Pero para poder llevar a cabo la competencia de pensamiento crítico, el estudiante necesita previamente trabajar las competencias de conocimiento y aplicación de problemas rutinarios.

Ante estos posibles efectos tan opuestos en el aprendizaje es importante que la implantación de la inteligencia artificial en las aulas se mida de la forma más objetiva y precisa posible. Este artículo trabaja justo esto mediante un estudio de caso. Primero, especificando cuál fue el objetivo en el uso de inteligencia artificial por parte de los alumnos, para más tarde poder cuantificarlo. Aunque, como se expondrá a lo largo del artículo, dado que se trata de un proyecto piloto en una fase inicial y no se ha llevado a cabo un experimento aleatorio, ni el uso de herramientas cuasiexperimentales, no hablaremos de relaciones causales.

Específicamente, este trabajo tiene por objetivo estudiar la relación entre el uso de la inteligencia artificial que fomenta la lectura del libro de texto de referencia (esto es *The Economy 1.0* de CORE Econ) y el aprendizaje de los estudiantes en la asignatura de Macroeconomía. Para ello vamos a analizar el estudio de un caso en el uso de inteligencia artificial en la Universidad Autónoma de Madrid, específicamente en el 2.º curso de la asignatura básica de Macroeconomía en dos grados diferentes, el grado de Economía y Finanzas (Econ&Fin) y el grado de Administración y Dirección de Empresas (ADE) pertenecientes a la Facultad de Ciencias Económicas y Empresariales. A lo largo de este trabajo, grados, grupos y clases se usan de forma similar, puesto que

en cada uno de los grados la profesora impartió a un único grupo y, por tanto, a una sola clase.

Para llevar a cabo este objetivo se han recopilado datos sobre las pruebas de evaluación en dos grupos diferentes impartiendo el mismo material por la misma profesora, pero con una diferencia importante: solamente uno de los grupos utilizó inteligencia artificial a través de la aplicación Honor Education (HE) (1). Estas pruebas forman parte de la evaluación continua y se desarrollaron en horario de clase. Comparando las mismas pruebas de evaluación para los dos grupos, concluimos que solamente en uno, de los dos exámenes parciales de razonamiento, el grupo que usó Honor Education tiene una nota media significativamente superior al grupo que no usó inteligencia artificial. En las tres pruebas que midieron conocimiento y aplicación de problemas rutinarios, los resultados no son tan claros, pues la nota media fue significativamente superior en el grupo que usó la inteligencia artificial para el test II, pero para el test III, ocurrió lo opuesto.

El trabajo aquí presentado pertenece a la literatura que analiza la importancia de los diferentes niveles competenciales cognitivos en el proceso de aprendizaje y cómo la inteligencia artificial podría potenciarlos. Estudios recientes demuestran que hay un interés en integrar la taxonomía de Bloom en la inteligencia artificial con objeto de mejorar el proceso de aprendizaje (2). Por ejemplo, Habiballa *et al.* (2025) comparan el rendimiento de las versiones 3.5 y 4.0 de ChatGPT con estudiantes humanos del área *Computer Science* categorizadas según la taxonomía de Bloom. Los resultados indican que, si bien los modelos de IA sobresalen en tareas de orden inferior como “Recordar” y “Aplicar”, los estudiantes humanos superan a la IA en tareas de orden superior que requieren “Comprender”, “Evaluar” y “Crear”. O como Scaria *et al.* (2024) cuya investigación evalúa la capacidad de los modelos de lenguaje de gran escala o *large language models* (LLM) para generar preguntas educativas en los distintos niveles de la taxonomía de Bloom. El estudio concluye que, con instrucciones adecuadas, los LLM pueden producir buenas preguntas en diferentes niveles competenciales cognitivos, aunque su efectividad varía entre modelos. Por el contrario, el

estudio de Bastani *et al.* (2024) investiga el efecto de GPT-4 en aproximadamente 1.000 estudiantes de secundaria en Matemáticas mediante un experimento aleatorio controlado usando dos sistemas de tutoría: GPT-Base (que proporciona respuestas directas) y una versión mejorada de este GPT-Tutor (que ofrece pistas paso a paso sin revelar directamente la solución completa). Los resultados demuestran que ambos estudiantes mejoran el rendimiento inmediato de las sesiones prácticas, pero que una vez que a los estudiantes les quitaron el asistente GPT, los estudiantes con GPT-Base disminuyeron su rendimiento y los de GPT-Tutor tuvieron el mismo resultado que el del grupo de control, que era el grupo que solamente utilizaba recursos educativos tradicionales.

El estudio anteriormente citado destaca una paradoja: aunque la IA generativa puede mejorar el rendimiento inmediato, puede también dificultar el desarrollo de habilidades fundamentales si se utilizan sin la orientación adecuada. Justo por ello, es importante entender cómo la IA puede potenciar los diferentes niveles cognitivos. Este trabajo analiza, aunque de forma preliminar, este efecto. Otros trabajos, a diferencia del aquí presentado, que se han llevado a cabo mediante diseños de experimentos aleatorios controlados (*RCT*, por sus siglas en inglés) proporcionando evidencia causal del impacto de las plataformas de IA en el aprendizaje son los de Banerjee *et al.* (2007) y Muralidharan *et al.* (2019). Ambos demuestran que las plataformas de aprendizaje adaptativo y personalizado y las intervenciones digitales dirigidas pueden generar mejoras significativas en el aprendizaje, especialmente en estudiantes con recursos limitados.

En lo que sigue, la sección segunda describe el caso del uso de inteligencia artificial en la Universidad Autónoma de Madrid. En esa sección se explica la plataforma de inteligencia artificial utilizada, y cuál fue su objetivo de aprendizaje. Además, se analiza la motivación de ese objetivo, así como el valor añadido del libro de CORE Econ para finalmente justificar los dos tipos de evaluaciones llevados a cabo durante el semestre. La sección tercera presenta los datos, comenzando con un análisis descriptivo de los dos grupos (el grado de Econ&Fin y el grado de ADE). La siguiente sección, la cuarta, presenta

los resultados sobre cómo se evaluó la adquisición de dichos objetivos. La sección quinta habla de las limitaciones del trabajo y debate sobre cómo mejorar de cara a futuro las limitaciones actuales del trabajo aquí presentado. Finalmente, la sección sexta concluye con un resumen de los resultados.

II. DESCRIPCIÓN DEL CASO

Al inicio del curso académico 2024-2025, CORE Econ y Honor Education (HE) —una plataforma educativa que utiliza la IA— se asociaron para ofrecer cursos introductorios de economía. Específicamente, el curso que se ha utilizado es el de Macroeconomía. En cursos anteriores en la Universidad Autónoma de Madrid, el material de CORE Econ ya lo había implantado en esta asignatura (3). Ahora se trataba de añadir AI mediante la plataforma HE. Se solicitó participar en el proyecto piloto de HE, por lo que la suscripción a la plataforma HE fue gratuita para los estudiantes pertenecientes al grado bilingüe de Economía y Finanzas (Eco&Fin), pero no fue utilizada en el grado de Administración y Dirección de Empresas (ADE). De esta forma he garantizado equidad en el acceso en el grado Eco&Fin de todos los alumnos matriculados en mi asignatura de Macroeconomía. Es bien sabido que una de las limitaciones de la IA es la falta de acceso equitativo a las herramientas de IA, que podría aumentar las desigualdades existentes entre estudiantes al tener que costearse el uso de las herramientas de pago.

Honor Education es una plataforma educativa impulsada por inteligencia artificial diseñada para mejorar la motivación de los estudiantes en su aprendizaje, ofrecer experiencias de aprendizaje personalizadas y garantizar la integridad académica. La plataforma aprovecha la inteligencia artificial y el aprendizaje automático para crear entornos de aprendizaje adaptativos, al tiempo que ofrece herramientas de evaluación, retroalimentación y comunicación, no solamente entre estudiantes, sino también entre estudiantes y profesores. Esto ayuda al profesor a identificar estudiantes con dificultades en el aprendizaje.

CORE Econ y Honor Education incluyen un curso en línea para estudiantes con actividades

de aprendizaje integradas para que los alumnos se familiaricen con el material antes y después de las sesiones presenciales, y guías para el docente de cada Unidad para ayudar en la planificación del curso y en las sesiones presenciales.

El curso introductorio de Macroeconomía se ha impartido en el primer semestre de este curso académico 2024-2025 en dos grados diferentes. En el grado de Economía y Finanzas (Econ&Fin), un grado bilingüe que se imparte en inglés, se ha utilizado la IA mediante el uso de la plataforma Honor Education, este sería el grupo tratado. En el grado de Administración y Dirección de Empresas (ADE) la asignatura era en castellano y no se ha utilizado Honor Education, este sería el grupo de control.

Las guías docentes de ambas asignaturas son prácticamente idénticas. Para ambas guías docentes se ha subido a Moodle (4) la correspondencia entre la guía docente de Macroeconomía de cada grupo y las secciones del libro de CORE Econ por unidades didácticas o temas. Como el contenido de las asignaturas en ambos grupos era prácticamente el mismo, la correspondencia entre la guía docente y las secciones del libro de CORE Econ eran las mismas. La gran diferencia entre ambos grupos ha sido que para el grupo tratado de Economía y Finanzas, se ha subido a Moodle el *link* a la plataforma Honor Education, con las secciones del libro de CORE Econ por capítulos. El libro de utilizado para este grupo ha sido *The Economy 1.0*. En el grupo de control de Administración y Dirección de Empresas no se ha utilizado Honor Education, y simplemente se dejaba en Moodle el *link* a las secciones del libro de CORE Econ *The Economy 1.0* traducido al español como *La Economía*. En ambos grupos se recomendaba la lectura a esos links antes (y después) de las clases presenciales. El contenido de los libros era el mismo, pero en el grupo de Economía y Finanzas utilizaban la inteligencia artificial y compartían información entre compañeros. En el grupo de Administración y Dirección de Empresas esto no era posible.

Por lo demás, las transparencias, al igual que las tareas, material complementario y los exámenes parciales, fueron idénticos para ambos grupos. Por tanto, alumnos que deciden ir a la misma universi-

dad y estudian en la misma facultad la misma asignatura, han compartido la misma profesora (con su misma metodología y estilo en clase).

A pesar de estas similitudes, hay que tener en cuenta que los dos grupos no son totalmente comparables, pues pertenecen a diferentes grados, y las notas de corte en el acceso a los grados difieren. De hecho, son siempre más altas en el grado de Economía y Finanzas que es el grupo bilingüe, aunque hay una tendencia a una disminución en las diferencias en la nota de corte. El único examen que fue diferente entre grupos fue el final, puesto que por un lado, en el grado de Administración y Dirección de Empresas el examen final es siempre coordinado entre los profesores que impartían la asignatura en ese grado, y por otro en la Universidad Autónoma de Madrid los exámenes finales se llevan a cabo en días diferentes según el grupo. Los resultados de ese examen final no han sido objeto de estudio aquí.

El objetivo principal que ha llevado a introducir la plataforma HE en el curso de Macroeconomía ha sido incentivar a los estudiantes la lectura del libro de CORE Econ como libro de texto de la asignatura. La era de la revolución digital, así como el cierre de los colegios durante el COVID-19, hizo que los profesores cada vez pongan más a disposición de los alumnos material de ayuda como resúmenes de los libros, transparencias, etc. Aunque a corto plazo esto ha sido una buena respuesta a las disrupciones generadas en el sector educativo debido a la pandemia, esto ha tenido un efecto indirecto negativo. Y es que ha reforzado una tendencia que venía ocurriendo entre la población estudiantil, pues cada vez menos estudiantes dedican tiempo a leer y reflexionar sobre el manual de la asignatura. Obviamente, las clases tienen por objetivo ayudar a pensar sobre los conceptos e ideas claves, pero la experiencia como docente durante todos estos años ha enseñado que para un aprendizaje completo, el alumno necesita dedicar tiempo individual en su estudio y practicar la concentración. Un buen libro de texto ayuda mucho en este proceso.

El hecho de que cada vez menos estudiantes universitarios dediquen tiempo a la lectura se está convirtiendo en un fenómeno global. En octubre

del 2024 (5), la revista *The Atlantic*, tras entrevistar a 33 profesores de diferentes universidades norteamericanas, enfatizaba que cada vez menos alumnos matriculados son capaces de leer un libro y parece que la razón detrás de esto es que tienen problemas para mantenerse concentrados. Leer libros, incluso por placer, no puede competir con TikTok, Instagram y Youtube. En España, también se está debatiendo sobre cómo está afectando este problema a la población adolescente (6).

El objetivo principal por el que se solicitó formar parte de la plataforma Honor Create Educatio utilizando la inteligencia artificial fue el de responder a la pregunta ¿los alumnos que dedicaban más tiempo a la lectura del libro de CORE Econ sacaban mejores resultados? ¿O, por el contrario, con las transparencias y ejercicios era suficiente? Para los que somos profesores, lamentablemente, esta última es una pregunta recurrente entre el alumnado. Y aunque la respuesta fue siempre “el libro”, nunca habíamos testado la hipótesis. La justificación de la respuesta recurrente es que el alumnado que no encuentra pasión por el conocimiento y el aprendizaje real busca como objetivo la calificación final y la obtención del título. Sin embargo, el aprendizaje que nos cambia la vida es aquel que perdura en el tiempo, trabaja el pensamiento crítico y la curiosidad por aprender. ¿Hasta qué punto un buen libro puede ayudar a adquirir este tipo de competencias en el aprendizaje?

Con este objetivo, el grupo de Economía y Finanzas tenía acceso gratuito a la plataforma HE con las secciones del libro de CORE Econ, en el otro grupo de Administración y Dirección de Empresas simplemente se adjuntó el *link* a las secciones del libro de CORE Econ en Moodle. Y para ambos grupos se enfatizó en clase la importancia de leer el manual.

Para poder valorar adecuadamente este impacto de la lectura del libro CORE Econ en el aprendizaje se necesitaba, por un lado, que la AI ayude y motive a los estudiantes en la lectura de los temas de CORE Econ, y por otro que las evaluaciones a los estudiantes fuesen consistentes con el valor añadido que el material de CORE Econ genera en el aprendizaje de los estudiantes.

Una característica clave de Honor Education es que permite la comunicación asincrónica con los estudiantes mediante anotaciones de texto y audio en archivos *epub* (sigla para *electronic publication*), vídeos y PDF. Las plantillas de diseño facilitan la visualización de las anotaciones integradas en la aplicación e incluyen marcadores de posición para que los estudiantes agreguen los suyos. La plataforma también incluye una rueda de reacción que permite a los estudiantes reaccionar al contenido e interactuar con el profesor y los demás alumnos de la clase compartiendo sus reacciones. Las reacciones de los estudiantes se guardan en sus cuadernos. Con los diferentes tipos de reacciones, el estudiante podía especificar, por ejemplo, si un párrafo o sección no estaba claro, o le ha parecido interesante o, por el contrario, le parece importante. Y estas reacciones, así como las notas, pueden ser públicas o privadas. Muchas de estas notas contenían dudas sobre el texto. Las respuestas del profesor a las dudas al estudiante pueden ser individuales, o bien visibles para toda la clase. Dado que este curso se lleva varios años impartiendo, y muchas eran preguntas que habían surgido otros años, las respuestas a esas preguntas se hacían visibles, sirviendo de herramienta para explicar dudas comunes a todos los estudiantes.

La hipótesis de este trabajo es entender si todas estas herramientas pertenecientes a Honor Education, y utilizadas únicamente en el grupo Economía y Finanzas, han motivado la lectura del libro, y por tanto, han mejorado el aprendizaje en este grupo más que en el grupo de Administración y Dirección de Empresas que no utilizó Honor Education.

Si esto es cierto cabría esperar que un posible mecanismo que lo genere es que, entre los estudiantes con menos tendencia a leer, los efectos de grupo —*peer effects*— sean mayores: al observar un estudiante que otros de la misma clase leen la sección, este se anima también a hacer lo mismo. De esta forma, la lectura de los capítulos, aunque es a nivel individual, al estar utilizando la aplicación HE, está incentivada por una colaboración activa (debido al uso de los botones de reacción y notas) y pública de los demás estudiantes en su lectura.

Puesto que para el grupo de Administración y Dirección de Empresas simplemente se añadió el link al libro de CORE Econ, pero sin poder utilizar las herramientas de reacción de la plataforma Honor Education, ni tampoco las notas públicas o privadas, este peer effect de la clase a través del uso de Honor Education, no ha podido ocurrir. Para estos estudiantes, se esperaba que hubiesen leído menos el libro, aunque sí habría podido ocurrir que estos alumnos leyeran por su cuenta el libro de CORE Econ a través del *link* en Moodle y comentasen entre ellos lo aprendido. Es importante destacar que, ante las preguntas en clase sobre si estaban leyendo el *link* al libro de CORE Econ, la respuesta del grupo de Administración y Dirección de Empresas era siempre negativa. Por lo tanto, sería de esperar que el *peer effect* fuese mucho más débil para los alumnos de Administración y Dirección de Empresas. Y que los alumnos del grupo de Economía y Finanzas hayan leído en media más capítulos del libro de CORE Econ que los del grupo de Administración y Dirección de Empresas.

El libro de CORE Econ se caracteriza por la renovación del temario tradicional de las enseñanzas de economía por uno orientado a una mejor comprensión de los problemas económicos actuales, con referencias a investigaciones recientes en economía, y donde se evitan prácticas docentes que lleven al alumno a una mera memorización de reglas y modelos teóricos aislados carentes de conexión con el análisis empírico y con la complejidad de la economía actual.

La mayoría de las pruebas de test estandarizados, y pruebas de conocimiento se centran en medir tres habilidades cognitivas: conocimiento, aplicación del conocimiento y razonamiento (7). El conocimiento analiza la capacidad de los estudiantes para recordar definiciones y hechos, y para reconocer características conocidas. La aplicación del conocimiento analiza la capacidad de los estudiantes para resolver problemas rutinarios, que normalmente se han practicado repetidamente en ejercicios de clase. Y el razonamiento evalúa la capacidad de los estudiantes para el pensamiento lógico y sistemático al confrontarlos con problemas complejos establecidos en con-

textos desconocidos. Las habilidades cognitivas de conocimiento y aplicación del conocimiento son a las que tradicionalmente se les ha dado más importancia en el sistema educativo, pero es la del razonamiento a la que cada vez se le está prestando más atención. El libro de CORE Econ ayuda mucho a que el estudiante trabaje esta última.

Tanto las actividades de evaluación como los criterios seguidos en la corrección fueron los mismos para ambos grupos. Las evaluaciones llevadas a cabo durante las catorce semanas del curso han consistido en tres exámenes de preguntas de opción múltiple así como dos exámenes parciales, llevándose a cabo todas ellas en horario de clase y los alumnos sabían con antelación las fechas o si el examen sería tipo test o de ejercicios. Mientras que los test de opción múltiple se centraban en preguntar sobre conceptos y problemas analizados en clase, los exámenes parciales se focalizaban en la aplicación de los conceptos y modelos vistos en clase al mundo real. Es decir, los exámenes parciales evalúan cómo usar los modelos para entender la actualidad económica, y, por tanto, la adquisición, del estudiante de competencias analíticas, pensamiento crítico y razonamiento.

III. EL ANÁLISIS DE LOS DOS GRUPOS

En esta sección analizamos, en primer lugar, información sobre los dos grupos de la Facultad de CC. Económicas y Empresariales pertenecientes a la Universidad Autónoma de Madrid, para pasar después a explicar el resultado de las pruebas en cada uno de los grupos, así como la comparación entre ambos.

El cuadro n.º 1 muestra algunos datos para ambos grupos. En el grupo tratado de Economía y Finanzas (Econ&Fin) para cada tema se añadía en Moodle el acceso a la plataforma Honor Education para que pudiesen leer las secciones del libro de CORE Econ y utilizar los botones de reacción. En el otro de Administración y Dirección de Empresas (ADE), simplemente se adjuntó para cada tema el *link* a las secciones del libro de CORE Econ en castellano. Y para ambos grupos se enfatizó en clase la importancia de leer el manual.

CUADRO N.º 1

DESCRIPTIVOS

	ECO & FINANZAS	ADE
Honor Education (HE)	Sí	No
Idioma	Inglés	Castellano
# Matriculados	72	64
% Mujeres	0,43	0,47
% Repetidores	0,17	0,13
% Erasmus	0,04	0
Nota corte 2023/2024	11,84	10,57
Nota corte 2024/2025	11,42	10,66
% de no presentados: evaluaciones		
Test I	0,07	0,29
Test II	0,07	0,29
Parcial I	0,09	0,07
Parcial II	0,12	0,04
Test III	0,12	0,05
% de no presentados: perfil		
Test nivel Matemáticas	0,19	0,25
Test nivel Economía	0,27	0,22

Fuente: Elaboración propia.

Como se puede observar, el grupo de Economía y Finanzas, comparado con el de Administración y Dirección de Empresas, se caracteriza por ser más numeroso y tener más estudiantes Erasmus, dado que las clases se imparten en inglés. El porcentaje de matriculados que repiten en la asignatura (17 por 100) es un poco mayor comparado con el de Administración y Dirección de Empresas (13 por 100). Dado que el curso pasado no se impartió en Administración y Dirección de Empresas, los repetidores de Economía y Finanzas eran antiguos estudiantes. En Economía y Finanzas, las mujeres representan solamente el 43 por 100, alcanzando casi el 50 por 100 (esto es 47 por 100) en el de Administración y Dirección de Empresas. Para el curso actual, la nota de Economía y Finanzas fue de 11,423 vs 10,667 (Administración y Dirección de Empresas) y la del curso académico pasado 11,843 versus 10,574.

El cuadro n.º 1 también muestra el porcentaje de alumnos no presentados a las pruebas de eva-

luación. Estas pruebas de evaluación, que son las mismas para ambos grupos, forman parte de la evaluación continua y todas las pruebas se desarrollaron en horario de clase. Ha habido un total de cinco pruebas de evaluación, tres exámenes tipo test y dos exámenes parciales. El orden de las evaluaciones por filas hace referencia al orden cronológico de las diferentes evaluaciones, con test de elección múltiple al inicio y al final del curso y en medio del semestre con dos parciales. Se puede observar que los alumnos de Economía y Finanzas en la última parte del curso (parcial II y test III) comienzan a ser más estratégicos y muchos de ellos dejan la asignatura para centrarse en otras, por lo que no se presentan a esas evaluaciones. Quizá la razón de esto es que estos alumnos saben que la asignatura de Macroeconomía no es de las fáciles de aprobar. Para el grupo de Administración y Dirección de Empresas pasa justo al contrario, y en las evaluaciones del final del curso se presenta un mayor porcentaje en comparación con el otro grupo.

Para conocer el nivel de cada grupo, al inicio del curso académico se prepararon unas preguntas de elección múltiple en horario de clase para hacer un análisis de perfil del grupo de conceptos básicos de matemáticas y de economía. Estas pruebas no formaban parte de la evaluación continua, solamente tenían como objetivo identificar alumnos con problemas, para, más tarde, escribirles un correo electrónico ofreciéndoles material de apoyo y animándoles a que asistiesen a tutorías individualizadas. El cuadro n.º 1 muestra que el porcentaje de no presentados era superior para ambos grupos en comparación con las pruebas que sí formaban parte de la evaluación continua.

IV. RESULTADOS

En esta sección explicamos los resultados de los exámenes comparando los dos grupos: el de Economía y Finanzas (bilingüe) y el de Administración y Dirección de Empresas.

El cuadro n.º 2 resume el contenido y las competencias evaluadas en los exámenes. Los exámenes tipo test sirven como método estandarizado

CUADRO N.º 2

CONTENIDO Y COMPETENCIAS DE LAS PRUEBAS DE EVALUACIÓN

	FECHAS DE EVALUACIÓN 2024	COMPETENCIAS EVALUADAS	CONTENIDO
Test I	25 y 27 septiembre	Definición y problemas rutinarios	PIB
Test II	2 y 4 octubre	Definición y problemas rutinarios	Inflación y paro
Parcial I	16 y 18 octubre	Razonamiento	Datos macro para explicar fluctuaciones económicas aplicadas al COVID-19
Parcial II	20 y 22 noviembre	Razonamiento	Modelo teórico/ para explicar políticas fiscales aplicado al <i>shock</i> DANA
Test III	18 y 20 diciembre	Definición y problemas rutinarios	IS-LM

Fuente: Elaboración propia.

CUADRO N.º 3

DIFERENCIAS EN EL ANÁLISIS DEL PERFIL

	GRADO ECONOMÍA & FINANZAS			GRADO ADMINISTRACIÓN Y D. EMPRESAS			DIFERENCIA ECONOMÍA & FINANZAS VS ADE			
	Media	Desviación estándar	N	Media	Desviación estándar	N	Dif. Media	Error estándar (Dif)	t	p-valor
Matemáticas	4,75	3,04	59	5,53	2,53	48	-0,78	0,54	-1,45	0,147
Economía	4,64	2,53	53	5,42	2,22	50	-0,79	0,47	-1,66	0,097

Fuente: Elaboración propia.

para evaluar de manera objetiva el conocimiento y las habilidades de reproducir problemas rutinarios hechos en clase. Los exámenes de preguntas con ejercicios, sin embargo, evaluaron la capacidad de razonamiento y pensamiento crítico. Estas preguntas consistían en llevar los modelos vistos en clase a la vida real. Cada pregunta de los ejercicios se evaluó siguiendo los mismos criterios en ambos grupos con objeto de ser lo más objetivo posible. Los alumnos sabían con antelación la fecha del examen, así como el tipo de examen (test o ejercicios) que se llevaría a cabo. Dado que ambos grupos hacían los mismos exámenes, y la asignatura no se impartía en el mismo día, el examen nunca se podía llevar a casa, y la corrección que se trabajó en clase se hizo una vez que el examen se había llevado a cabo en ambos grupos.

Queremos testar si existe evidencia estadísticamente significativa de que la nota media de las pruebas de evaluación difiere entre el grupo de Economía y Finanzas, que utilizó Honor Education, y el grupo de Administración y Dirección de Empresas, que no utilizó la inteligencia artificial en la lectura del libro CORE Econ. La hipótesis nula es que las medias poblacionales de las notas de los exámenes son iguales en ambos grupos, frente a la alternativa bilateral de que las medias difieren. Si los alumnos de Economía y Finanzas han dedicado más tiempo a leer el libro, se esperaría que estuviesen más preparados en aquellos exámenes que midan la capacidad de razonar. En los exámenes tipo test que miden conocimiento y aplicación de problemas rutinarios en el examen, a ambos grupos se esperaría que tuviesen medias en notas similares (8).

CUADRO N.º 4

DIFERENCIAS EN LAS EVALUACIONES DE LOS ESTUDIANTES

	GRADO ECONOMÍA & FINANZAS			GRADO ADMINISTRACIÓN Y D. EMPRESAS			DIFERENCIA ECONOMÍA & FINANZAS VS. ADE			
	Media	Desviación estándar	N	Media	Desviación estándar	N	Media	Error estándar (Diff)	t	p-valor
Test I	3,37	2,28	67	4,16	2,44	46	-0,79	0,46	-1,72	0,0854
Test II	6,95	2,47	67	6,04	2,11	46	0,91	0,44	2,07	0,0384
Parcial I	3,77	2,03	66	4,05	1,98	60	-0,28	0,36	-0,78	0,4354
Parcial II	5,38	2,32	64	4,29	2,34	62	1,19	0,42	2,83	0,0094
Test III	4,53	2,72	64	5,69	2,15	61	-1,16	0,44	-2,64	0,0082

Fuente: Elaboración propia.

Los cuadros n.º 3 y 4 muestran la media, la desviación estándar y el número de alumnos que hicieron en clase el test en cada grupo. Además, nos muestra información sobre la diferencia de medias del grupo de Economía y Finanzas menos el grupo de Administración y Dirección de Empresas, así como los errores estándares de estas diferencias, el estadístico *t* y el *p*-valor. Lo primero que llama la atención es que hay mucha variabilidad en las notas, independientemente del grupo analizado, de si son pruebas sobre el análisis del perfil (cuadro n.º 3) o pruebas de evaluación (cuadro n.º 4). Esto explica, en parte, que los errores estándares de las diferencias en las medias muestrales sean, en muchos casos, elevados, y la estimación de la diferencia en medias sea poco precisa. Los tamaños muestrales de cada una de las pruebas, al ser mayor de 30, son razonables, por lo que el estadístico *t* presenta una distribución normal estándar. El cuadro n.º 3 muestra que en las pruebas que no forman parte de la evaluación continua, y que sirven para detectar el nivel de la clase al comienzo del curso escolar, no se encuentran diferencias significativas entre ambos grupos, ni en Matemáticas, ni en Economía.

Los resultados de los exámenes que forman parte de la evaluación continua vienen dados en el cuadro n.º 4 y sugieren varias conclusiones. En primer lugar, no siempre el grupo que utilizó Honor Education obtiene mejores resultados. De hecho, las diferencias en medias entre el grupo que utilizó la aplicación Honor Education (es decir, Economía

y Finanzas) y el que no la utilizó (es decir Administración y Dirección de Empresas) varían desde un -0,79 puntos el valor mínimo a unos 1,19 puntos el máximo. Segundo, esas diferencias son estadísticamente significativas al 1 por 100 solamente en dos de los cinco exámenes. Específicamente, esto ocurre en el Parcial II donde el grupo que utilizó Honor Education obtiene 1,19 puntos en media superiores al que no la utilizó, y en el Test III, donde el grupo de Administración y Dirección de Empresas que no utilizó Honor Education sacó ventaja 1,16 puntos en media. Los 1,9 puntos superiores para el grupo de Economía y Finanzas podrían estar relacionados con una mayor adquisición de competencias de razonamiento y aplicación de los modelos teóricos a la vida real con la lectura del libro de CORE Econ. El desarrollo de estas competencias podrían haber sucedido en la plataforma Honor Education debido a los *peer effects* mediante *social learning* o del *competition effect*. Específicamente, en el primer caso, los estudiantes de ese grupo ven que sus compañeros de clase leen más y ellos deciden leer también (*social learning*). En el segundo caso, los estudiantes terminan leyendo más en esta clase debido a que es una clase muy competitiva (una característica que suele definir al grupo de Economía y Finanzas). Y tercero, de los tres exámenes tipo test, las diferencias estimadas en las notas son significativas en dos de ellos (el test II y III), pero no son estables, pues en el test II el grupo que utilizó Honor Education sacó mejores notas, pero en el test III ocurre exactamente lo contrario.

Por lo que, en lo que respecta al conocimiento y la aplicación de problemas rutinarios, no siempre el grupo de Economía y Finanzas obtuvo mejores resultados. Este análisis no nos cuenta por qué podrían existir estas diferencias. La siguiente sección habla un poco sobre ello.

V. LIMITACIONES

Una de las limitaciones de este trabajo es el problema de selección debido a que alumnos diferentes *ex ante* deciden estudiar grados diferentes. A pesar de que las pruebas de nivel que se hizo al alumnado justo al inicio de curso no concluyen que haya diferencias significativas en la media de las notas, probablemente los alumnos de Economía y Finanzas, al ser un grupo bilingüe, provengan de familias con diferente nivel socioeconómico y cultural. Para resolverlo, necesitaríamos llevar a cabo idealmente un experimento aleatorio o bien usar técnicas cuasiexperimentales.

Otra limitación del trabajo aquí presentado es que no identifica ni testa los mecanismos a través del cual los *peer effect* aparecen en clase. La evidencia empírica nos enseña que los efectos de las notas y del bagaje socioeconómico de los compañeros de clase son importantes en magnitud (Hoxby, 2000; Hoxby y Weingarth, 2005; Ammermueller y Pischke, 2006) y que pueden actuar de forma no lineal. Es decir, no todos los alumnos se benefician por igual del nivel académico del resto. Tal como comenta Sacerdote (2011), el principal reto empírico en el análisis de las interacciones sociales tiene que ver con la identificación de sus verdaderos efectos, ya que su estimación recoge diversos problemas. Por un lado, la existencia de dos fuentes de endogeneidad: el llamado *reflection problem* (Manski, 1993), que surge porque las notas de un alumno influyen en las de sus compañeros y viceversa, y el problema de *sorting*, explicado en el párrafo anterior, que surge porque la selección de alumnos en clases no se realiza de forma aleatoria. Por otro, las dificultades empíricas para identificar por separado la influencia de las notas de los compañeros de la influencia del bagaje socioeconómico. Adicionalmente, aparece la dificultad de separar el efecto de la calidad del profesor del efecto de los compañeros, puesto que

el profesor es un factor que afecta al conjunto del aula, que es el marco en el que se producen las interacciones entre los estudiantes. Por ello, es crucial observar al mismo profesor en diferentes clases (Rockoff, 2004), algo no muy frecuente en los datos, pero que sí ocurre en este trabajo.

VI. CONCLUSIONES

- El objetivo de este estudio es analizar la posible relación entre el uso de la inteligencia artificial, que incentiva la lectura del libro de CORE Econ (libro de referencia de la asignatura), y el aprendizaje del estudiante.
- Para poder llevar a cabo este análisis es fundamental disponer tanto de información sobre el rendimiento académico de los estudiantes en el que un mismo profesor imparta clases en diferentes grupos de la misma asignatura como el que solamente en uno de los grupos se haya utilizado la inteligencia artificial.
- En la plataforma Honor Education los alumnos tienen acceso a los botones de reacción con la lectura del libro de CORE Econ, pudiendo esas reacciones hacerlas públicas para toda la clase. En este contexto, se esperaría que aparezcan nuevos efectos de los compañeros de clase en la plataforma Honor Education.
- Las cinco pruebas de evaluación, que son las mismas y utilizan los mismos criterios para evaluar, miden no solamente conocimientos y aplicación de problemas rutinarios trabajados en clase a través de tres test de elección múltiple, sino también cómo aplicar esos conocimientos a problemas reales y la capacidad de razonamiento, a través de dos exámenes parciales de ejercicios. Esta mezcla de evaluaciones intentan captar el valor añadido del libro de texto de CORE Econ.
- Los resultados demuestran que, en uno de los dos exámenes parciales de ejercicios, el grupo expuesto a la IA para incentivar la lectura del libro de CORE Econ sacó una media significativamente mayor en 1,19 puntos, en compara-

ción con el otro grupo. Por el contrario, de las tres evaluaciones tipo test de elección múltiple, en dos de ellas las diferencias de medias en las notas son significativas. En el test II la media es mayor para el grupo que utiliza la inteligencia artificial, pero en el test III ocurre lo contrario.

Estas conclusiones nos enseñan que, si queremos diseñar políticas educativas que mejoren el rendimiento escolar, un paso previo es entender la manera en la que la inteligencia artificial afecta el proceso de aprendizaje.

NOTAS

- (1) Para un análisis más detallado del valor añadido de la plataforma de IA en el sector educativo HE, ver sección segunda.
- (2) La taxonomía de Bloom (Bloom's Taxonomy) es una teoría pedagógica que analiza diferentes niveles de competencias cognitivas que los estudiantes deberían alcanzar en el proceso de aprendizaje, que comienzan por el conocimiento, pasando por aplicar y analizar y termina en la capacidad de crear y evaluar.
- (3) CORE Econ es un proyecto que crea recursos de enseñanza de economía de libre acceso. Para más detalle ver <https://www.COREEcon.org>.
- (4) Moodle es la plataforma que utilizan los docentes de la Universidad Autónoma de Madrid para la creación de sus propios sitios web privados con el objetivo de colgar el material relacionado con la asignatura impartida.
- (5) Para una lectura del artículo ver: <https://www.theatlantic.com/magazine/archive/2024/11/the-elite-college-students-who-cant-read-books/679945/>
- (6) Ver noticia en *El País* del 7 de abril del 2025 <https://elpais.com/mamas-papas/expertos/2025-04-08/beatriz-gimenez-de-ory-escritora-los-adolescentes-han-cambiado-el-no-me-gusta-leer-por-el-no-puedo-es-muy-largo.html>

- (7) Ver, por ejemplo, los test estandarizados de PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) o NAEP (National Assessment of Educational Progress).
- (8) Diferentes estudios demuestran que los exámenes tipo test también miden la capacidad estratégica de tomar decisiones ante el riesgo, ver, por ejemplo, Baldiga (2014). Los test aquí presentados en ambos grupos penalizaban (-0,33), pudiendo crear distorsiones en la evaluación si no todos los candidatos asignan la misma probabilidad a que cualquiera de las opciones sea correcta o difieren en actitud ante el riesgo.

BIBLIOGRAFÍA

- Ammermueller, A., y Pischke, J-S. (2006).** Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, vol. 27(3), 315-348.1264.
- Baldiga, K. (2017).** Gender Differences in Willness to Gues. *Management Scienc* 60(2), 434-448.
- Banerjee, A., Cole, S. Duflo, E. y Linden, L. (2007).** Remediating education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235- 1264.
- Bastani, H., Bastani O., Sungu, A., G. E., H. Kabakc, O. y Mariman, R. (2024).** Generative AI Can Harm Learning. *Wharton School Research Paper*, July 2024 .
- Habiballa, H., Kotyrba, M., Volna, E., Bradac, V., Dusek, M. (2025).** Artificial Intelligence (ChatGPT) and Bloom's Taxonomy in Theoretical Computer Science Education. *Appl. Sci.*,15, 581.
- Hanushek, E. y Kimko, D. (2000).** Schooling, labor force quality, and the growth of nations. *American Economic Review*, 90, nO. 5 (December), pp 1184-1208.
- Hanushek, E. A. y Woessmann, L. (2012).** Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, December 2012, 17(4), 267-321

- Hoxby, C. (2000).** Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Paper*, No. 7867. https://www.nber.org/system/files/working_papers/w7867/w7867.pdf
- Hoxby, C. y Weingarth, G. (2005).** Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. *NBER Conference Paper*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=386de4e83235738ec333387e23abf71a4c1f1e4d>
- Manski, C. (1993).** Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60, 531-542.
- Muralldharan, K., Singh, A. y Ganimina, A. J. (2019).** Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4), 1426-1460.
- Murnane, R. J., Willet, J. B. y Levy, F. (1995).** The Growing Importance of Cognitive Skills in Wage Determination. *Review of Economics and Statistics*, 77, 251-266.
- Rivkin, S. G., Hanushek, E. A. y Kain, J. F. (2005).** Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417-458.
- Sacerdote, B. (2011).** Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? *Handbook of Economics of Education*, vol. 3, capítulo 4.
- Scaria, N., Chenna, S. D. y Subramani, D. (2024).** Automated Educational Question Generation at Different Bloom's Skill Levels using Large Language Models: Strategies and Evaluation. *Artificial Intelligence in Education*.
- Vojtech, M., Havel, H. y Vojtechová, E. (2025).** Artificial Intelligence (ChatGPT) and Bloom's Taxonomy in Theoretical Computer Science Education. *Applied Sciences*, 15(2), 581, 1-34.