La penalización del aprendizaje por la IA generativa: evidencia de la educación secundaria china

Resumen traducido en base al estudio

de David Strömberg (Universidad de Estocolmo), Victor Lei y Yanhui Wu (Universidad de Hong Kong). Junio de 2026.

La pregunta del estudio.

La inteligencia artificial generativa (herramientas como ChatGPT, DeepSeek o Doubao) permite hacer tareas mucho mas rápido. Pero en el colegio el objetivo no es terminar la tarea, sino **aprender haciéndola**. El estudio se pregunta que pasa con el aprendizaje real de los estudiantes cuando usan IA por su cuenta, en condiciones normales y durante periodo largo.

Cómo se hizo.

- **Quiénes:** 26.811 estudiantes de 7°a 12° grado (equivalente a 7°básico a IV° medio en Chile) de nueve colegios de un condado del centro de China. Cubre el 90% de los estudiantes de secundaria de la zona.
- **Cuanto tiempo:** 30 meses, de septiembre de 2022 a junio de 2025.
- **Qué se midió:**
  1. Notas y tiempo dedicado a las tareas semanales, registrados en una plataforma digital.
  2. Pruebas mensuales sin apuntes ni ayuda, hechas en clase.
  3. Exámenes de admisión de alto impacto: el **Zhongkao** (para entrar a la enseñanza media y el **Gaokao** (para entra a la universidad).
- **Uso de la IA:** una encuesta de junio de 2025 pregunto a cada estudiante en que mes empezó a usar IA y cuantas horas por semana lo usa. Los profesores les pidieron revisar la fecha de registro en sus aplicaciones para mayor precisión.

El método.

Como los estudiantes empezaron a usar IA en distintos momentos, los investigadores compararon como cambiaron las notas de cada estudiante antes y después de usarla, frente a estudiantes similares que nunca la usaron. Antes de adoptar la IA, ambos grupos tenían notas practicas idénticas y evolucionaban igual. Por eso las diferencias posteriores pueden atribuirse con bastante confianza a la IA.

Contexto.

El uso de IA paso de casi cero en 2022 a cerca del 80% de los estudiantes en junio de 2025. Las herramientas las usadas fueron Doubao (47%) y DeepSeek (36%). Las materias donde mas se usa son matemática (66%) e inglés (55%).

Principales resultados.

1. **Mejores tareas, pero peores pruebas.** Seis meses después de empezar a usar IA:
   - Las **notas de las tareas suben un 18%.**
   - El **tiempo para hacerlas baja de 64 a 45 minutos** (cerca de un 30% menos).
   - Las **notas de las pruebas mensuales caen**

El efecto negativo crece de forma gradual durante unos seis meses y luego se estabiliza. Una explicación es que los estudiantes tardan en aprender a usar bien la IA para hacer las tareas por ellos.

2. **El daño en los exámenes de admisión es igual de grande, pero tarda más en aparecer**
   Para quienes llevaban dos años o mas usando IA, la caída fue del **24% en el Zhongkao** y del **18% en el Gaokao**. Estos exámenes evalúan lo aprendido durante varios años, así es que el deterioro se acumula lentamente. Esto sugiere que los estudios de corta duración, que son la mayoría, **subestiman el costo real a largo plazo**.

3. **El problema es "externalizar" la tarea**
   Cerca del 81% de usuarios de IA (tras cinco meses de uso) muestran un patrón de dejar que la IA haga la tarea por ellos:
   - Terminan la tarea en menos tiempo que el mas rápido de los estudiantes que no usan IA (menos de 50 minutos).
   - Sacan notas muy altas en las tareas, similares a lo que logra la propia IA.
   - Les va muy mal en las pruebas sin ayuda.

El hallazgo más esperanzador es que los usuarios de IA que siguen dedicando a la tarea el mismo tiempo que los demás (50 a 65 minutos) tienen notas en las pruebas similares a quienes no usan IA. La IA en sí no les quita capacidad de aprender. Lo que daña el aprendizaje es dejar de hacer el esfuerzo.

4. **Materias más afectadas**
   - **Ciencias sociales** (Política, Geografía, Historia): -27% en promedio.
   - **Ciencias y matemática** (Física, Química, Biología, Matemática): - 22%
   - **Inglés** -17%
   - **Chino** (lengua materna): -9%

5. **Estudiantes más afectados**
   Los investigadores dividieron a los estudiantes en grupo y midieron la caída en las pruebas mensuales entre 6 y 10 meses después de empezar a usar IA. En los exámenes de admisión, los patrones fueron muy parecidos.

   - **Los mas jóvenes:** en 7° a 9° grado la caída es del 24%, frente al 17% en 10° a 12° grado. Según los profesores, los mayores tienen mas trabajo supervisado y mas restricciones al uso de IA.

   - **Los que más la usan:** quienes usan IA 0 a 1 hora semanal pierden un 5%, mientras que quienes la usan 5 horas o mas pierden alrededor de un 30%. El daño crece mas rápido al principio y se estanca a partir de unas 4 horas mensuales.

   - **Los hombres:** pierden un 21,6% frente al 18,4% de las mujeres. Esto se explica casi completamente porque ellos la usan más horas.

- **Los mejores alumnos:** el tercio con mejores notas pierde un 24%, frente al 16% del tercio con notas mas bajas. Esto reduce la distancia entre buenos y malos estudiantes, pero porque los mejores caen más, no porque los más débiles mejoren.

6. Señales de adaptación

El daño tras cinco meses de uso, los puntajes en las pruebas mensuales sin ayuda bajaron cerca del 25% a inicio de 2023 a cerca del 16% en junio de 2025. Este patrón también aparece al seguir siempre al mismo grupo de estudiantes. Esto sugiere que estudiantes o profesores se están adaptando, aunque la perdida siendo considerable.
Esto significa que, con el tiempo, usar IA sigue bajando las notas, pero menos al principio.

¿Por qué nadie reaccionó antes?

- Para el **colegio o condado**, el efecto total parecía pequeño al principio: solo un 3,4% de caída promedio hasta junio de 2023. Llego al 9,9% en junio 2025, cuando la mayoría ya usaba IA.
- **Cada profesor ve solo su materia**, y una caída fuerte en una sola asignatura no es tan rara.
- **Los estudiantes no perciben que aprenden menos**. Además, la caída es gradual y es difícil relacionarla con la IA.
- **Las tareas pierden su valor como señal**. Una buena nota en la tarea ya no indica que el estudiante aprendió. Entre usuarios de IA, incluso puede indicar lo contrario.

Comprobaciones de solidez

Los autores verificaron sus resultados de varias formas:

- Otros métodos estadísticos dieron resultados casi idénticos.
- Las **pruebas comunales**, con las mismas preguntas para todos los colegios, mostraron la misma caída (-23%) esto descarta que los profesores hayan cambiado la dificultad de sus pruebas.
- **No hubo efecto en los compañeros** que no usan IA.
- **Caídas tan grandes son casi inexistentes** entre estudiantes que no usan IA, así es que es poco probable que se deben a otras causas, como enfermedades o problemas personales.

Relación con otros estudios

- **Coincide con un experimento en Turquía,** donde el acceso a una IA que daba respuestas directas en matemática mejoro las tareas, pero bajo los resultados en las pruebas en un 17%.
- **No contradice un estudio en Harvard,** donde un tutor de IA bien diseñado mejoró el aprendizaje de física. La diferencia es que, en la vida real, la mayoría de los adolescentes elige herramientas que dan respuestas rápidas y no tutores pedagógicos, aunque estos existen y suelen ser gratuitos.

Recomendaciones de los autores

1. **Informar** a estudiantes y familias sobre el costo a largo plazo de dejar que la IA haga las tareas.
2. **Dar mas peso a las evaluaciones presenciales sin ayuda,** para que el esfuerzo vuelva a tener recompensa.
3. **Que padres y profesores supervisen el esfuerzo y el tiempo dedicado,** y no solo las notas de las tareas, que ya no reflejan bien lo aprendido.

En síntesis

La IA generativa hace que las tareas salgan mejor y mas rápido: las notas de las tareas suben un 18% y el tiempo para hacerlas baja de 64 a45 minutos. Sin embargo, cerca de 8 de cada 10 estudiantes que usan IA la aprovechan para que haga la tarea por ellos, saltándose el esfuerzo de resolver los ejercicios. Ese esfuerzo es justamente lo que produce el aprendizaje. Por eso, aunque la tarea se vea bien, lo aprendido cae con fuerza: en las pruebas sin ayuda, la nota baja alrededor de un 20% en unos seis meses.

El daño no se queda en las pruebas del mes. En los exámenes de admisión a la enseñanza media y a la universidad, que evaluan lo aprendido durante varios años, la perdida se va acumulando y alcanza su nivel completo despues de unos dos años de uso: alrededor de un 24% en el Zhongkao y un 18% en el Gaokao. Esto significa que los efectos mas graves pueden tardar en notarse. Tambien significa que estudios de pocas semanas o meses probablemente subestiman el problema.

No todos pierden lo mismo. Los más afectados son los estudiantes más jóvenes, los que usan IA más horas a la semana, los hombres y, sorprendentemente, los mejores alumnos. Por materia, el golpe es mayor en ciencias sociales, seguido de ciencias y matemáticas, y es menor en idiomas.

La buena noticia es que el problema no está en la IA en sí, sino en cómo se usa. Los estudiantes que usan IA, pero siguen dedicando a la tarea el mismo tiempo que los demás obtienen notas en las pruebas similares a las de quienes no la usan. Además, el daño ha ido disminuyendo con el tiempo, lo que sugiere que estudiantes y profesores están empezando a adaptarse. Aun así, la pérdida sigue siendo considerable.

Para los autores, la clave está en los incentivos. Proponen informar a estudiantes y familias sobre el costo de dejar que la IA haga las tareas, dar más peso a las evaluaciones presenciales y sin ayuda, y que padres y profesores se fijen en el esfuerzo y el tiempo dedicado más que en las notas de las tareas, que ya no reflejan bien lo que el estudiante realmente sabe.

ASTORECA
EDUCAR BIEN