

Evaluación del efecto de la inteligencia artificial en el aprendizaje

EDICIÓN BREVE

PISA 2025 · Deberes · Tutores automáticos · Motivación · Evidencia experimental ·
España · Pautas

Índice general

Antes de empezar	1
1 El examen sin la máquina	1
2 PISA 2025 de cerca	4
3 Peor que hace quince años, no que hace cien	8
4 El alumno que no quiere	9
5 Los ensayos, uno por uno	11
6 Milagros y marchas atrás	14
7 El profesor y la evaluación	15
8 España	16
9 Cómo se hace aprender a quien no quiere	17
10 Qué hacer	18
Cómo leer los veredictos	19
Referencias	20

Antes de empezar

Sobre la IA en la escuela se oyen dos cosas. Que cada niño tendrá por fin el tutor particular que antes solo tenían los ricos, y que una generación entera está dejando de pensar. Las dos las sostiene gente seria, y las dos tienen ya datos que las apoyan. Quien vende la herramienta promete una revolución; quien la teme cita el peor estudio; y de educación se habla casi siempre en tono de derrumbe, porque cada generación ha creído que la siguiente se echaba a perder.

Puede que esta vez sea distinto. Este informe no lo decide por simpatía: lo mide y lo razona. Mira qué ha pasado con las notas, qué hacen de verdad los alumnos cuando tienen un tutor automático delante y qué ensayos se han hecho bien. Y toma en serio una sospecha incómoda: que el alumno medio no quiere aprender más de lo imprescindible. Los datos la matizan y, en lo esencial, la confirman.

CAPÍTULO 1

El examen sin la máquina

—30 %

tiempo de deberes de 26.811 alumnos chinos de secundaria a los seis meses de empezar a usar IA: de 64 a 45 minutos. Su nota de deberes sube un 18 %¹

—20 %

nota de esos mismos alumnos en el examen mensual a libro cerrado, a los seis meses; —24 % y —18 % en los exámenes de acceso, donde la pérdida tarda unos dos años en completarse¹

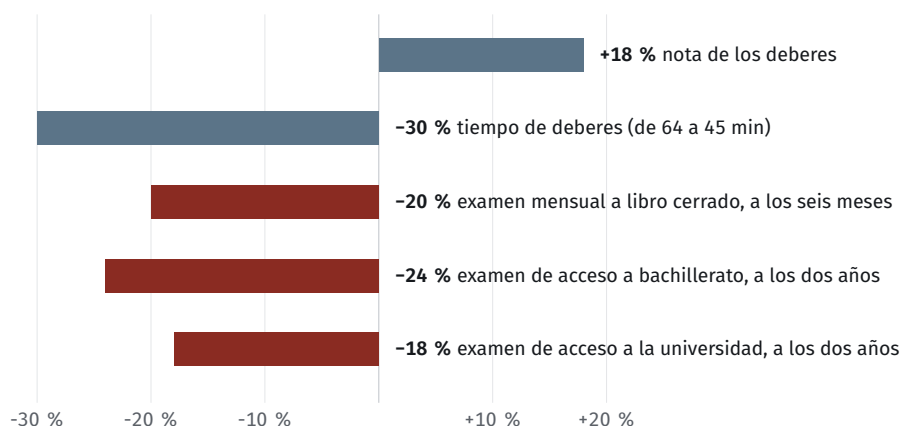
4 de cada 5

usuarios que terminan los deberes más deprisa que el más rápido de los que no usan IA¹

La IA ha vuelto gratuita e inagotable la explicación, que en la escuela nunca fue lo escaso. Lo escaso es un motivo para sentarse, y eso no lo produce. De ahí sale todo este informe, empezando por su dato más duro.

Treinta meses en un condado chino. El mejor dato sobre lo que hace la IA con un alumno corriente lo obtuvieron Strömberg, Lei y Wu, tres economistas de Estocolmo y Hong Kong: los registros de los 26.811 alumnos de 7.º a 12.º de un condado del centro de China, de 2022 a 2025: duración y nota de los deberes de cada semana, exámenes mensuales a libro cerrado y los dos exámenes de acceso. Al empezar casi nadie usaba IA generativa; al acabar, el 80 %, con asistentes generales sin ninguna función educativa. Como cada alumno empezó en una fecha distinta, se compara a quien empieza con quien aún no ha empezado; no hay sorteo.¹

Figura 1 Lo que cambia cuando un alumno empieza a usar IA generativa



Strömberg, Lei y Wu (2026), 26.811 alumnos de 7.º a 12.º de un condado chino, 2022-2025. Porcentajes sobre la media anterior a la adopción.

En el examen mensual, la pérdida está completa a los seis meses; en los de acceso tarda unos dos años. Los estudios de pocas semanas, avisan los autores, pueden subestimar el coste. Pierden más los pequeños, los chicos y, sobre todo, los que antes iban mejor. En el trabajo, la IA iguala porque sube a los de abajo; en la escuela, porque baja a los de arriba.

No daña la herramienta, sino el uso que elige la mayoría. Cuatro de cada cinco usuarios acaban antes que nadie, con notas de deberes que coinciden con lo que sabe hacer un chatbot, y con exámenes peores. El quinto sigue dedicando el mismo tiempo que antes y saca en el examen lo mismo que quien no usa IA. Existen tutores bien diseñados y gratuitos,

dicen los autores, y «la mayoría de los alumnos sencillamente no los usa»; el problema «no desaparecerá con los avances de la técnica».

El mismo resultado, medido por dentro. ALEKS, una plataforma estadounidense de matemáticas, guarda desde 2015 cuánto tarda cada alumno en cada problema. Desde finales de 2022, el tiempo en los problemas de enunciado —que se pegan en un chatbot— cae respecto a los de gráficas, que no se pueden pegar: nada en 5.º de primaria, un 9 % en los primeros cursos de secundaria, un 31 % en el instituto. Con un vigilante delante, el tiempo vuelve a ser el de siempre, y en esas pruebas vigiladas la razón de probabilidades de acertar ha bajado un 25 %: para quien partiera de un 80 % de aciertos, quedarse en un 75 %.² Los autores lo llaman rendición: pegar el enunciado y copiar la respuesta se salta el esfuerzo de recordar, que es lo que fija lo aprendido.

Los ensayos con sorteo dicen lo mismo. En un instituto turco, practicar con un asistente sin restricciones dio un 17 % menos en el examen posterior sin ayuda; con uno que solo daba pistas no hubo daño, y tampoco mejora.³ La OCDE lo resume así: la ventaja de quien usa herramientas generales «desaparece, y a veces se invierte, en los exámenes».⁴

Se rompe el termómetro. Entre los alumnos chinos sin IA, mejor nota en los deberes significa mejor examen. Entre los que la usan, la relación se invierte. En una universidad israelí con 36.000 alumnos, las notas de las asignaturas con trabajos en casa suben, los suspensos caen un tercio, y el puesto de un alumno en ellas predice cada vez peor su puesto en las de examen presencial.⁵ El psicólogo Donald Campbell lo advirtió en 1979: cuanto más se usa un indicador para decidir, más se corrompe.⁶ La nota de los deberes servía porque hacerlos costaba lo mismo que aprender; desde que los hace una máquina, mide a la máquina. Hipótesis de este informe: el primer daño de la IA no es al aprendizaje, sino a la medida. Hasta que llega un examen sin máquina, todo parece ir mejor que nunca.

Por qué el alumno no lo nota. En un ensayo de Harvard, los alumnos creyeron aprender más en la clase magistral pulida y sacaron mejor nota tras la sesión de trabajo, que les resultó incómoda.⁷ El psicólogo Robert Bjork lleva treinta años documentándolo: lo que hace fluida la práctica mejora el rendimiento del momento y empeora lo que queda.⁸ Un chatbot es la explicación más fluida jamás inventada: leer su solución produce la sensación exacta de haberla entendido.

Por qué estos veredictos. Los dos primeros ya se miden, en dos países y con métodos distintos, y su causa no es técnica: mientras acabar antes sea posible y nadie compruebe después, la mayoría acabará antes. El tercero es «muy improbable» por lo que enseña el quinto alumno: quien mantiene el tiempo de trabajo aprende lo mismo. Lo que se pierde es práctica, y la práctica se recupera.

VEREDICTO Hasta 2031

Las tareas hechas en casa dejan de servir como medida de lo que sabe un alumno

■■■■■ **Muy probable**

Grave para la evaluación; ya ocurre

La mayoría de los alumnos que usan IA por su cuenta la usan para acabar antes, no para entender

■■■■■ **Muy probable**

Grave y silencioso: solo se ve en un examen sin máquina

Una generación «incapaz de pensar» por culpa de la IA

■ ■ ■ ■ ■ **Muy improbable**

Catastrófico; nada de lo medido apunta ahí

Escala de los veredictos, siempre de aquí a 2031. Se explica al final del informe.

CAPÍTULO 2

PISA 2025 de cerca

—23

puntos de España en lectura entre 2022 y 2025, hasta 451; —16 en matemáticas (457) y —7 en ciencias (477). Desde 2015, —45 en lectura: más de dos cursos^{9,10}

0

veces que aparecen «IA», «inteligencia artificial» o «ChatGPT» en el recuadro con el que la OCDE explica por qué ha caído la lectura¹¹

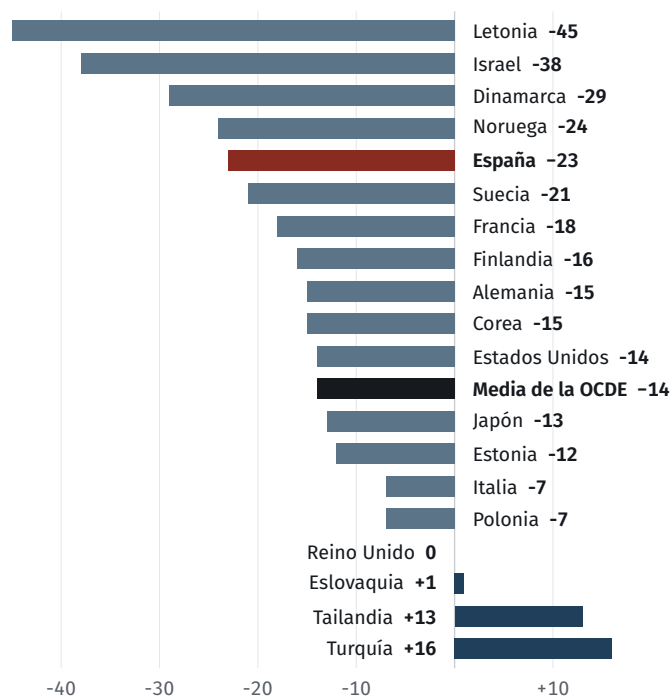
—11 %

tiempo diario de deberes que declaran los alumnos de la OCDE respecto a 2022; baja en casi todos los países a la vez¹²

El peor resultado de la serie, en casi todas partes. La media de la

OCDE es la más baja desde que existe la prueba: 482 en ciencias, 461 en lectura y 463 en matemáticas. España queda por debajo en las tres. Y el dato real puede ser algo peor: las exclusiones de alumnos subieron al 7,9 %, con un 23,2 % en Cataluña, cuyos resultados no se publican, y la OCDE avisa de que eso sesga al alza.¹³

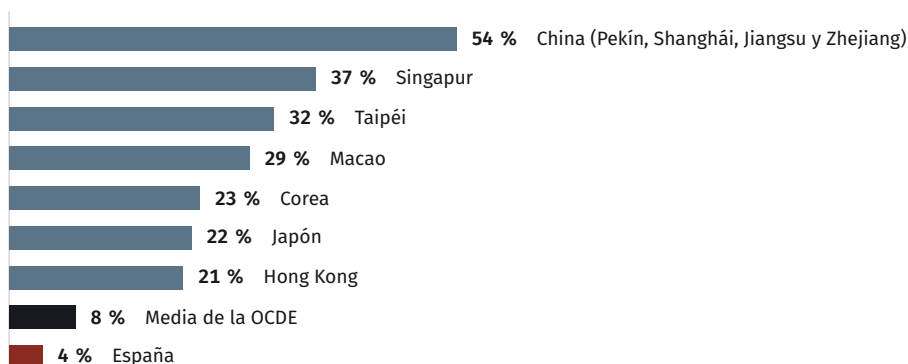
Figura 2 Variación de la puntuación en lectura entre 2022 y 2025



OCDE, PISA 2025, volumen I, tabla I.1. Selección de países.

El problema español está arriba. España no tiene más alumnos por debajo del mínimo que la media. Tiene la mitad de alumnos excelentes: 4 % en ciencias, 4 % en matemáticas y 2 % en lectura, frente al 7, 8 y 6 %. Y en España como en la OCDE, la caída reciente es mayor entre los alumnos favorecidos: comprime desde arriba, igual que en el condado chino.¹³

Lo que la OCDE dice que ha pasado. Cuatro fuerzas, y ninguna es la IA. El alumno aguanta menos: lo que más cae son los textos largos; los lectores «apresurados», que responden rápido y mal, pasan del 7 al 11 %;

Figura 3 Alumnos excelentes en matemáticas: niveles 5 y 6 de PISA 2025

OCDE, PISA 2025, nota de país de España.

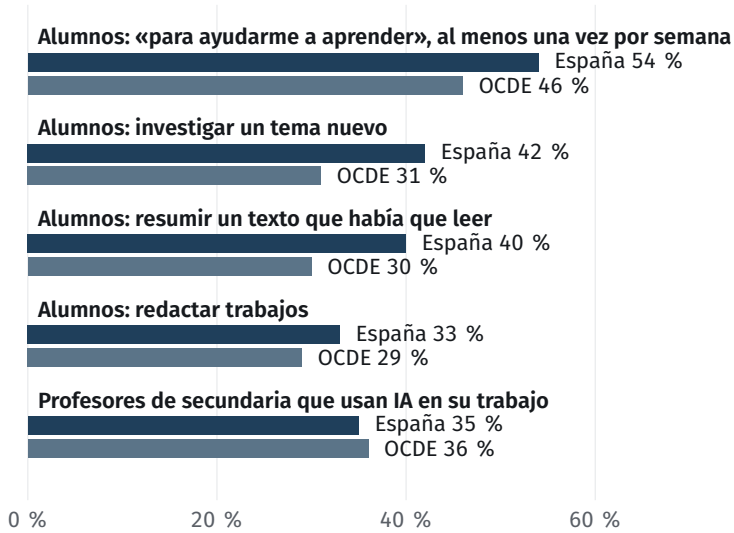
se rinde peor al final de la prueba que al principio; se declara menos esfuerzo. El aula es más difícil: la parte de la clase que se va en mantener el orden sube del 13 al 16 %. El sistema responde tarde a grupos más heterogéneos. Y la sociedad lee menos, también los padres. La lectura cae desde 2012 y las matemáticas desde 2018.¹¹

Dos ideas cómodas que el informe no respalda. El tiempo de pantalla fuera de la escuela bajó entre 2022 y 2025, mientras caían las notas. Y España ya ha prohibido el móvil: el 86 % de sus alumnos va a centros donde está vetado, frente al 67 % en 2022 y al 49 % de la OCDE, y ha perdido 23 puntos en lectura. Eso no dice qué habría pasado sin prohibirlo; dice que no bastó.^{12, 13}

Qué pinta la IA. El 54 % de los alumnos españoles usa chatbots para estudiar al menos una vez por semana, frente al 46 % de la OCDE; solo el 8 % no los usa nunca. Quien no usa IA para resumir, redactar o buscar puntúa más que quien la usa; es una asociación: puede reflejar quién recurre a la herramienta tanto como el efecto de usarla. Y el tiempo de deberes cae un 11 % en tres años en casi todos los países, los mismos tres años de ChatGPT.¹²

Lo firme es que la caída empezó antes. Cuánto añade la IA, PISA no lo puede medir, y que la OCDE no la nombre tampoco prueba que no pese. Hipótesis de este informe: la encontró hecha y la acelera por la pieza que ya fallaba, el esfuerzo sostenido, allí donde nadie mira: en casa. Los examinados en 2025 llevaban dos cursos y medio con chatbot. Los del

Figura 4 Uso de chatbots de IA al menos una vez por semana, y uso de IA por el profesorado



OCDE, PISA 2025, nota de país de España (alumnos de 15 años); OCDE, TALIS 2024 (profesorado de secundaria).

próximo PISA, en 2029, habrán hecho con él toda la secundaria.

Quién aguanta. El Reino Unido no pierde nada en lectura; los nórdicos, los sistemas más digitalizados de Europa, se hunden: Dinamarca –29, Noruega –24, Suecia –21.⁹

Por qué estos veredictos. Las fuerzas que describe la OCDE no han cambiado de signo, y quienes se examinen en 2029 habrán hecho toda la secundaria con un chatbot; no es «muy probable» porque el Reino Unido demuestra que se puede no caer. Lo segundo lo descarta la cronología: la lectura baja desde 2012.

VEREDICTO Hasta 2031

PISA 2029 con resultados iguales o peores que los de 2025 en la media de la OCDE

La IA, y no otra cosa, como causa principal de la caída medida en 2025

■ ■ ■ ■ ■ Probable

Grave; cuarta caída seguida

■ ■ ■ ■ ■ Muy improbable

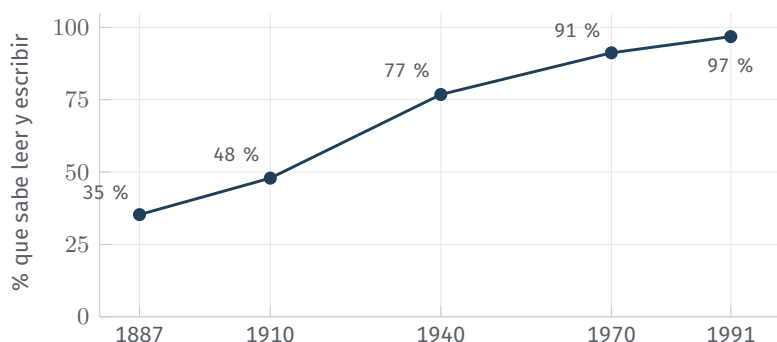
La caída empieza diez años antes

CAPÍTULO 3

Peor que hace quince años, no que hace cien

«Mínimo histórico» significa mínimo desde el año 2000. En 1910 sabía leer y escribir el 48 % de los españoles mayores de diez años; en 1970, el 91 %. ¹⁴ La escuela obligatoria hasta los 16 años es de 1990. Los estadounidenses de 13 años leían en 2023 como en 1971 y siete puntos peor que en 2012, y en 2025 siguen igual. ^{15,16} En Noruega, las puntuaciones de inteligencia bajan desde los nacidos en 1975, también entre hermanos: es el ambiente, y empezó antes del móvil. ¹⁷ No es la generación más ignorante de la historia. En varias pruebas y países rinde peor que la de hace quince años, después de un siglo de avances enormes. Es grave porque interrumpe una mejora muy larga, y recuperable por lo mismo.

Figura 5 Población española de diez o más años que sabe leer y escribir



N. de Gabriel (1997), a partir de los censos de población. Hasta 1910, población de once o más años.

Fuera de los países ricos, el problema es otro. Hay 272 millones de niños y jóvenes sin escolarizar, harán falta 44 millones de profesores más de aquí a 2030 y, en los países de renta baja y media, siete de cada diez niños de diez años no entienden un texto sencillo. ^{18,19,20} Allí el signo de la IA cambia: la misma herramienta que en un país rico sustituye esfuerzo, en uno pobre sustituye ausencia.

CAPÍTULO 4

El alumno que no quiere

17 %

sesiones con errores en las que el alumno mediano escribió algo al tutor de IA, casi siempre una respuesta suelta o un clic, en un ensayo de dos cursos en 18 institutos de Tennessee; el 96 % lo había probado²¹

60

sistemas educativos donde ha subido desde 2022 la parte de alumnos a quienes aprender cosas nuevas les aburre; no ha bajado en ninguno⁹

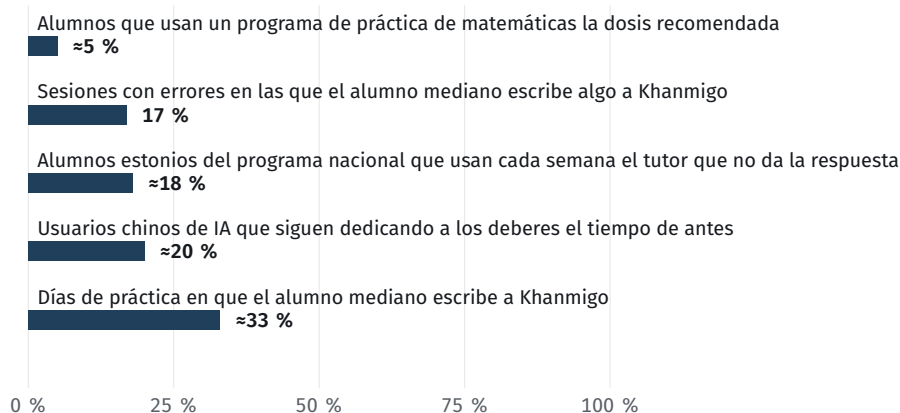
47 %

hogares españoles con hijos en edad escolar que pagan clases particulares: 2.782 millones de euros al año²²

Dos cursos con el tutor a un clic. Alumnos flojos de 18 institutos de Tennessee, asignados al azar, en horario lectivo, a una plataforma de práctica con un tutor de IA que orienta y no da la respuesta. El programa sirvió: entre 0,06 y 0,08 desviaciones típicas al año. Al tutor es difícil atribuirle nada, porque apenas se usó: el alumno mediano le escribió uno de cada tres días, y casi siempre una respuesta suelta o un clic. Conclusión de los autores: la promesa exige conseguir que los alumnos lo usen, no darles acceso.²¹ El fundador de la plataforma reconoció que para la mayoría el tutor había sido un «no acontecimiento».²³

El patrón es anterior a la IA y aparece en los cuatro programas con datos. A quien cumple la dosis de los programas de práctica en línea le va mejor —es una asociación: quien la cumple no es un alumno cualquiera—, y la cumple el 5 %.²⁴ En Estonia, el mejor país de Europa en PISA, el asistente nacional que pregunta en vez de responder lo usa cada semana menos de uno de cada cinco destinatarios, cuando antes del programa entre el 64 y el 90 % ya usaba herramientas comerciales.²⁵ La herramienta que obliga a pensar compite con la que da la respuesta, y el alumno ya tiene la segunda en el bolsillo.

Lo que dicen de sí mismos y lo que hacen. Los alumnos españoles declaran más perseverancia que la media de la OCDE y obtienen el peor resultado de su serie. El alumno medio de la OCDE declara 1,4 horas

Figura 6 Cinco medidas del uso real, cada una con su denominador

Holt (2024); Oreopoulos y Low (2026); Fundación TI-Hüpe (marzo de 2026); Strömberg, Lei y Wu (2026). Medidas y poblaciones distintas; ninguna mide directamente el esfuerzo. Muestran problemas de adopción y de continuidad.

diarias de deberes, no cero; pero es tiempo declarado, y ha bajado un 11 % en tres años.^{12, 13}

Lo que compran las familias. Una explicación clara de cualquier tema de secundaria es hoy gratuita, instantánea e inagotable, y las familias españolas siguen pagando casi tres mil millones al año en clases particulares. En ESO y Bachillerato la participación ha bajado cinco puntos desde 2019; hipótesis de este informe: el chatbot se come la parte del negocio que consistía en explicar, que era la barata.²² La encuesta no pregunta por qué se paga. La lectura de este informe es que, en el refuerzo escolar, se paga sobre todo un adulto que espera al chico dos tardes por semana y se da cuenta si no ha hecho nada.

De ahí la tesis central de este informe, que es una interpretación y no el resultado de un estudio: en el aprendizaje de un adolescente corriente, lo escaso nunca fue la explicación. Es la obligación. La IA ha vuelto infinita la oferta de lo que sobraba y no produce nada de lo que falta. No es pereza moral: estudiar cuesta hoy y rinde dentro de dos años, y casi todo el mundo pesa de más el presente. En 2002, Dan Ariely y Klaus Wertenbroch lo midieron con universitarios: los plazos impuestos desde fuera rindieron más que los que cada uno se ponía.²⁶ Una academia o un examen el jueves son eso: un compromiso que otro hace cumplir. Un

chatbot está siempre disponible y nunca espera a nadie.

Por qué estos veredictos. El primero ya se mide en cuatro programas. El segundo es «improbable» porque el ensayo más largo da casi nada con un tutor que casi nadie usó, y nada en el diseño voluntario corrige el sesgo hacia el presente.

VEREDICTO Hasta 2031

La mayoría de los alumnos con acceso libre a un tutor de IA lo usa poco o para obtener la respuesta

■■■■■ **Muy probable**

Grave: anula el beneficio

Un tutor de IA de uso voluntario mejora de forma apreciable al alumno medio de un país rico

■■■■■ **Improbable**

Los ensayos largos dan entre nada y muy poco

CAPÍTULO 5

Los ensayos, uno por uno

552.000

accesos al metaanálisis que atribuía a ChatGPT un efecto «grande» sobre el rendimiento, retractado el 22 de abril de 2026²⁷

0,37

desviaciones típicas: efecto medio de la tutoría humana en 96 ensayos con sorteo²⁸

0,06–0,08

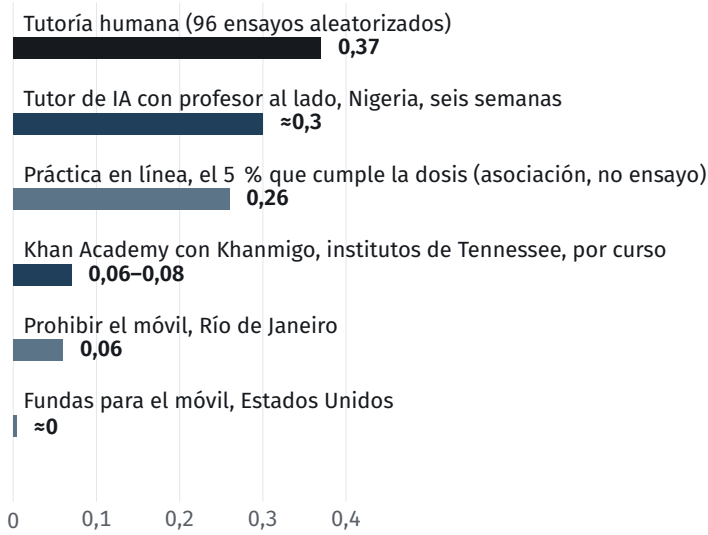
efecto anual de la plataforma con tutor de IA en el ensayo más largo hecho en una escuela corriente²¹

La promesa y la vara de medir. En 1984, el psicólogo Benjamin Bloom publicó que un alumno con tutor particular rendía dos desviaciones típicas más que en clase: es la cifra que citan quienes venden tutores automáticos.²⁹ Noventa y seis ensayos con sorteo la han dejado en 0,37.

Un artículo retirado. El metaanálisis que se citaba para decir que la ciencia avalaba el chatbot en el aula —51 estudios, efecto grande, 433 citas— fue retractado por «discrepancias» que minan sus conclusiones;

los autores no respondieron.²⁷ Siguen publicándose otros con efectos de esa talla, hechos con estudios de pocas semanas que miden con la herramienta delante.

Figura 7 Tamaño del efecto sobre el aprendizaje, en desviaciones típicas



Nickow, Oreopoulos y Quan (2020); De Simone et al. (2025); Holt (2024); Oreopoulos y Low (2026); Lichand et al. (2026); Allcott et al. (2026). Poblaciones, pruebas y diseños distintos: la barra de práctica en línea es una asociación; las demás, estimaciones causales.

Tabla 1 Los ensayos principales sobre IA generativa y aprendizaje, 2024-2026

Estudio	Resultado	Pega principal
Turquía, ~1.000 alumnos de instituto ³	Asistente libre: –17 % en el examen sin ayuda. Con pistas: ni daño ni mejora	Pocas sesiones
Harvard, Física, 194 alumnos ³⁰	Más del doble de ganancia que en clase	Dos lecciones, prueba inmediata
Nigeria, seis semanas por las tardes ³¹	≈0,3 d. t. en papel; mejora el examen de fin de curso	Lo evalúa quien lo promueve

continúa

Estudio	Resultado	Pega principal
EE. UU., la IA sugiere al tutor humano ³²	+4 puntos de lecciones superadas; +9 con tutores flojos	Sin mejora detectable a fin de curso
Tennessee, dos cursos ²¹	0,06–0,08 d. t.; el tutor apenas se usó	No separa tutor y plataforma
OpenAI, «modo estudio» ³³	Microeconomía ≈ 15 % más; neurociencia, nada	Lo hace el fabricante; examen inmediato

La mejora vive en la medida cercana. Puestos en fila, los estudios dibujan algo que ninguno dice por separado: cuanto más cerca del uso se mide, mejor sale. Ejercicios resueltos durante la práctica, tiques de salida, el examen del mismo día. Donde se mide lejos y sin máquina, el efecto es nulo, pequeño o negativo, con una excepción: Nigeria, donde había adulto y horario y mejora también el examen de fin de curso. Es una hipótesis: los estudios comparan intervenciones distintas, no un mismo tratamiento medido a varias distancias. El propio OpenAI admite que lo que importa es si las ganancias «siguen ahí con el tiempo», y que aún no lo sabe.³³

Donde sí ha funcionado. En Nigeria había un sitio, una hora, profesores presentes y una prueba final en papel, y la alternativa no era un profesor particular sino nadie.³¹ En el ensayo de Stanford, la IA no habla con el alumno: le sopla al tutor humano qué preguntar —la herramienta cuesta veinte dólares por tutor y año—, y los tutores flojos alcanzan a los buenos.³²

Cinco preguntas para el próximo titular. ¿Se midió sin la herramienta? ¿Cuánto tiempo después? ¿Quién decidía usarla? ¿Comparado con qué? ¿Quién lo hace y quién lo paga?

CAPÍTULO 6

Milagros y marchas atrás

7

alumnos por escuela, de media, en la prueba con la que El Salvador anunció resultados «como los de Alemania y Suecia»: 1.198 voluntarios de 171 centros³⁴

346

puntuación de El Salvador en matemáticas en el PISA de verdad, con 7.073 alumnos: tres puntos más que en 2022 y 117 menos que la media de la OCDE³⁵

≈0

efecto medio sobre las notas de guardar el móvil en una funda con cierre, en el mayor estudio hecho en Estados Unidos³⁶

El Salvador. La prueba no fue PISA, sino una herramienta que la OCDE ofrece a escuelas sueltas. Siete voluntarios por escuela. Las escuelas entraron en el piloto, según el propio Gobierno, por su infraestructura, su conexión y su disposición. Sin medida anterior ni grupo de comparación. Y el piloto no era un chatbot: incluía dispositivos, método de enseñanza, materiales y acompañamiento al profesorado. La ministra y el Banco Mundial advirtieron de que los resultados valen solo para esas escuelas.³⁴ Lo notable es que El Salvador tiene un dato bueno que nadie usa: en el examen nacional sube doce puntos en ciencias mientras escolariza a más jóvenes, antes del acuerdo con xAI y sin relación demostrada con ningún chatbot.^{11,35}

Corea gastó unos 850 millones de dólares en libros de texto con IA; la adopción cayó del 37 al 19 % en un semestre y el Parlamento les quitó el rango de libro de texto a los cuatro meses.³⁷ **China**, que es quien más empuja la IA, prohíbe desde mayo de 2025 a los alumnos de primaria usar por su cuenta chatbots abiertos y veta entregar como deberes contenido generado.³⁸ Llegó cuando cuatro de cada cinco alumnos del condado del primer capítulo ya los usaban en casa. También pide a las familias que supervisen; nadie ha medido si eso cambia lo que pasa en un dormitorio.

La bala de plata anterior. El mayor estudio sobre prohibir el móvil encuentra un efecto sobre las notas «sistemáticamente cercano a cero»; en Río de Janeiro, 0,06 desviaciones típicas.^{36,39,40} Entre nada y un poco,

según el sitio y el tiempo: no era la palanca. Con el chatbot hay una razón añadida para no esperar mucho de prohibirlo en clase: el daño medido ocurre en casa.

Por qué estos veredictos. Los anuncios sin grupo de comparación salen gratis: el titular llega el primer día y la corrección, meses después. Lo segundo es «improbable» porque solo Estonia ha encargado la evaluación que permitiría demostrarlo.

VEREDICTO Hasta 2031	
Más anuncios de resultados espectaculares de programas nacionales de IA sin grupo de comparación	 Muy probable Limitado; desorienta
Un programa nacional de IA educativa con mejora demostrada en un examen externo	 Improbable Beneficioso; Estonia es la candidata

CAPÍTULO 7

El profesor y la evaluación

Alumnos que la usan cada semana, profesores que dicen no estar formados. En 2024 usaba IA en su trabajo el 35 % de los docentes españoles de secundaria, sobre todo para preparar clases; tres de cada cuatro de los que no la usaban decían no tener formación. Un año después, el 54 % de los alumnos la usa cada semana para estudiar.^{13,41} Son encuestas distintas, pero lo que consta basta: quien tiene que rediseñar los deberes y los exámenes dice no estar preparado para ello.

Lo que cambia no es la explicación, es la evaluación. Los autores del estudio chino recomiendan mirar el tiempo que se dedica a los deberes, que dice más que su nota, y comprobar lo trabajado en casa con pruebas cortas en clase.¹ La OCDE propone un orden: primero sin IA, después con IA educativa, después con IA general.⁴ Es el de la calculadora, que no perjudicó las destrezas cuando entró después de aprender a mano.⁴² Con una diferencia: la calculadora se queda con la aritmética y deja el

planteamiento; el chatbot se queda con todo.

Hipótesis de este informe: el trabajo del profesor de secundaria se desplaza de explicar a hacer trabajar y comprobar. Explicar lo hace ya cualquier asistente. Ninguno consigue que treinta adolescentes pasen cincuenta minutos intentando algo difícil, ni sabe cuál no ha entendido.

CAPÍTULO 8

España

2.818 M€

coste presupuestario neto anual, de media entre 2027 y 2036, de la ley que baja la ratio a 22 alumnos en primaria y 25 en ESO⁴³

2.782 M€

lo que gastan al año las familias españolas en clases particulares²²

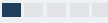
La apuesta es bajar la ratio. Costará 5.169 millones en 2031, el 13,8 % del gasto público en primaria y ESO, y es regresiva: las aulas llenas están en barrios acomodados. El metaanálisis más completo, con 66 estudios causales, no encuentra un efecto medio significativo de reducir la clase sobre el aprendizaje. Los autores del cálculo recomiendan llevar el dinero a lo que sí lo tiene: tutoría en grupo pequeño.⁴³ El Estado va a gastar en una medida sin efecto medio demostrado lo mismo que los hogares gastan, por su cuenta, en algo parecido a lo que sí lo tiene.

Donde la IA tendría sentido. Detrás de un tutor, en horario escolar: es donde hay efectos medidos, y la herramienta cuesta veinte dólares al año por tutor; lo caro son los tutores. Nadie lo ha propuesto. El debate español se reduce a si se permite el chatbot en clase, que es donde menos daño hace.

Los que llegan ahora. Los examinados en el próximo PISA, el de 2029, habrán tenido chatbot desde 4.º de primaria. Lo razonable es esperar que las notas internas suban, los suspensos bajen y los exámenes externos empeoren. España tiene pocos: la prueba de acceso, las evaluaciones de

diagnóstico y PISA. Son los únicos termómetros que la IA no ha roto.

Por qué estos veredictos. Lo primero ya se ha medido en China y en Israel, y España tiene pocos exámenes externos que lo delaten. Lo segundo: el anteproyecto está presentado y el mejor metaanálisis no encuentra efecto medio. Lo tercero serían 45 puntos de lectura en cuatro años; ningún país ha recuperado eso en un ciclo.

VEREDICTO Hasta 2031	
Suben las notas internas y bajan las de las pruebas externas	 Probable Grave; oculta el problema
Ley de ratios aprobada, sin mejora apreciable en evaluaciones externas	 Probable Grave para el presupuesto
España recupera en PISA 2029 el nivel de 2015	 Muy improbable —

CAPÍTULO 9

Cómo se hace aprender a quien no quiere

Ningún chatbot ha demostrado motivar a quien no quiere. Lo que hace trabajar a un adolescente sin ganas es lo de siempre, y el diseño que funciona es el que no depende de que quiera.

Uno: la obligación la pone una persona. La tutoría humana rinde 0,37 desviaciones típicas, y más dentro del horario escolar que después.²⁸ El metaanálisis no aísla por qué. La lectura de este informe: pesa menos la explicación, que también da un vídeo, que el hecho de que alguien ha quedado con el alumno, le mira trabajar y vuelve el jueves.

Dos: el examen sin máquina, corto y frecuente. Intentar recordar sin mirar enseña más que releer: es el experimento clásico de Roediger y Karpicke.⁴⁴ Diez minutos en papel cada semana sobre los deberes de esa semana obligan a recuperar, devuelven al profesor el termómetro y convierten los deberes copiados en un mal negocio.

Tres: mirar el tiempo, no la nota. Si una hoja que llevaba una hora se entrega perfecta en quince minutos, ha pasado algo.¹

Cuatro: primero sin, luego con. Se aprende a hacerlo a mano antes de delegarlo, porque solo quien sabe hacerlo distingue un resultado bueno de uno malo.

Cinco: si hay tutor de IA, que no espere a que le llamen. Integrado en el ejercicio, interviene ante el error y no deja avanzar sin acertar: los alumnos van más despacio y rinden algo mejor una semana después.²³ Con alumnos corrientes, que el tiempo por problema se desplome es mala señal.

Seis: la IA, detrás del adulto. Preparar materiales, proponer al tutor la siguiente pregunta, corregir un primer borrador. Es donde están los efectos medidos.³²

Lo que no va a pasar. No aparecerá un asistente tan bueno que el alumno sin ganas lo prefiera al que le da la respuesta: cuanto mejor sea el modelo, mejor hará los deberes. Y no bastará con enseñar «uso responsable»: saber lo que conviene nunca ha bastado a los quince años.

CAPÍTULO 10

Qué hacer

Tres conclusiones para unos años. La IA no originó la caída de las notas, que empezó antes; lo más probable es que la acelere por los deberes. Sobran buenas explicaciones; falta un motivo para sentarse, y eso lo dan una persona, un horario y un examen. Y la nota de lo hecho en casa ha dejado de informar: toda decisión sobre un alumno necesita una medida tomada sin máquina.

Si estudia. Una semana después de trabajar un tema con IA, intente un ejercicio parecido sin nada delante: si no sale, no lo sabe. Separe las sesiones de producir de las de aprender. Pida que le examine o que encuentre el fallo, no la respuesta. Si con la IA tarda mucho menos, desconfíe y compruébelo. Busque ante quién responder: un compañero, una academia, un preparador, una fecha. Y lea largo, en papel, sin teléfono: es lo que más ha caído y lo que más falta hace.

Si tiene hijos. Interésese por lo que aprenden cada día, no por el boletín. Mire el tiempo: una tarea difícil terminada en un cuarto de hora

y sin un tachón no es buena noticia. Pídale que le explique un ejercicio. El teléfono, fuera del cuarto de estudio; en primaria, sin asistentes abiertos. Si paga clases particulares, sepa qué compra: la explicación ya es gratis; lo que el teléfono no regala es un adulto que espera y comprueba.

Si enseña. Califique lo que ha visto hacer: pruebas cortas y frecuentes en papel, preguntas orales, trabajos que se defienden de palabra. Los deberes sirven para practicar, no para poner nota. Primero a mano. Use la IA para usted. Y enseñe a dudar de la máquina con casos concretos.

Si decide sobre un presupuesto. Tutoría en grupo pequeño dentro del horario antes que ratios para todos. Ningún despliegue sin grupo de comparación ni examen sin la herramienta. Y conservar y publicar las evaluaciones externas.

Lo que queda por encima de todo esto. Quien ha aprendido algo difícil sabe de sí mismo una cosa que no se obtiene de otro modo: que puede. La mayoría lo descubre en la escuela, casi siempre porque alguien le obligó a quedarse un rato más con un problema del que quería huir. Una máquina que resuelve el problema le ahorra el mal rato y le quita el descubrimiento. El maestro, el padre que pregunta y el compañero de estudio no son sustitutos imperfectos de un tutor automático: son la parte que nunca fue un problema de información. Y donde esa persona no existe —harán falta 44 millones de profesores más de aquí a 2030—, un asistente en el teléfono puede ser lo primero que se parece a un maestro. Las dos cosas son verdad a la vez.

Cómo leer los veredictos

Los veredictos usan siempre la misma escala y el mismo plazo, de aquí a 2031. Son juicios de este informe a partir de los datos citados; no son frecuencias medidas. En este informe, varios escenarios son favorables: la segunda línea dice entonces qué se ganaría.



Referencias

Numeradas por orden de primera cita. El informe extenso detalla el alcance y los límites de cada fuente. Consulta: septiembre de 2026.

- 1 Strömberg, D., Lei, V. y Wu, Y.. *The Generative AI Learning Penalty: Evidence from Chinese Secondary Education*. 2026-06-02. cepr.org ↗
- 2 Rismanchian, S., Uzun, H., Matayoshi, J., Cosyn, E. y Kurd-Misto, E.. *Faster Completion, Less Learning: Generative AI Reduced Study Time on Math Problems and the Knowledge They Build*. 2026-06. arxiv.org ↗
- 3 Bastani et al.. *Generative AI without guardrails can harm learning*. 25-06-2025. doi.org ↗
- 4 OCDE. *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*. 2026-01-19. oecd.org ↗
- 5 Hausman, N., Rigbi, O. y Weisburd, S.. *Generative AI in universities: Grades up, signals down, skills in flux*. 2025-06-17. cepr.org ↗
- 6 D. T. Campbell. *Assessing the impact of planned social change*. 1979. doi.org ↗
- 7 Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K. y Kestin, G.. *Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom*. 2019. pnas.org ↗
- 8 N. C. Soderstrom y R. A. Bjork. *Learning Versus Performance: An Integrative Review*. 2015. doi.org ↗
- 9 OCDE. *PISA 2025 Results (Volume I): Future-Ready Students. Executive summary y tabla I.1*. 2026-09-08. oecd.org ↗
- 10 Ministerio de Educación, Formación Profesional y Deportes. *El alumnado español reporta niveles de perseverancia ante el aprendizaje superiores al promedio de los países de la OCDE, según el informe PISA 2025*. 2026-09-08. lamoncloa.gob.es ↗
- 11 OCDE. *PISA 2025 Results (Volume I), capítulo 2: Student performance in PISA 2025 (recuadro I.2.3, «Why has reading performance declined over time?»)*. 2026-09-08. oecd.org ↗
- 12 OCDE. *PISA 2025 Results (Volume I), capítulo 4: Student school life and beyond*. 2026-09-08. oecd.org ↗
- 13 OCDE. *PISA 2025 Results (Volume I). Country notes: Spain*. 2026-09-08. oecd.org ↗
- 14 De Gabriel, N.. *Alfabetización, semialfabetización y analfabetismo en España (1860-1991)*. 1997. revistas.ucm.es ↗
- 15 National Center for Education Statistics. *NAEP Long-Term Trend Assessment Results: Reading and Mathematics, age 13*. 2023-06. nationsreportcard.gov ↗

- 16 National Center for Education Statistics. *NAEP Long-Term Trend Assessment Results 2025*. 2026. nationsreportcard.gov ↗
- 17 Bratsberg, B. y Rogeberg, O.. *Flynn effect and its reversal are both environmentally caused*. 2018. pnas.org ↗
- 18 UNESCO. *Countries commit to reducing the number of out-of-school children by 165 million*. 2025. unesco.org ↗
- 19 UNESCO y Equipo Especial Internacional sobre Docentes. *Global Report on Teachers*. 2024. unesco.org ↗
- 20 Banco Mundial, UNICEF, UNESCO y otros. *70 % of 10-Year-Olds now in Learning Poverty, Unable to Read and Understand a Simple Text*. 2022-06-23. worldbank.org ↗
- 21 Oreopoulos, P. y Low, N.. *One Click Away: AI Tutoring with Khanmigo in a Two-Year School Experiment*. 2026-08. nber.org ↗
- 22 Moreno, J. M. y Gortazar, L.. *El mercado de las clases particulares en España entre 2020 y 2024*. 2026-06-25. esade.edu ↗
- 23 Chalkbeat. *Students rarely engaged with Khan Academy's AI-powered tutor Khanmigo, study finds*. 2026-08-25. chalkbeat.org ↗
- 24 Holt, L.. *The 5 Percent Problem: Online mathematics programs may benefit most the kids who need it least*. 2024. educationnext.org ↗
- 25 TI-Hüpe (AI Leap Foundation). *Estonia Launches Nationwide AI Leap Education Program*. 2026. tihupe.ee ↗
- 26 D. Ariely y K. Wertenbroch. *Procrastination, Deadlines, and Performance: Self-Control by Precommitment*. 2002. doi.org ↗
- 27 Humanities and Social Sciences Communications (Springer Nature). *Retraction Note: The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis*. 2026-04-22. doi.org ↗
- 28 Nickow, A., Oreopoulos, P. y Quan, V.. *The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence*. 2020-07. nber.org ↗
- 29 B. S. Bloom. *The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring*. 1984. doi.org ↗
- 30 Kestin et al. (Harvard). *AI tutoring outperforms in-class active learning*. 2025-06. nature.com ↗
- 31 De Simone, M. E., Tiberti, F., Mosuro, W., Manolio, F., Barron, M. y Dikoru, E.. *From chalkboards to chatbots: Transforming learning in Nigeria, one prompt at a time*. 2025-01-09. blogs.worldbank.org ↗
- 32 Wang et al. (Stanford). *Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise*. 2024-10. nssa.stanford.edu ↗
- 33 OpenAI. *New tools for understanding AI and learning outcomes*. 2026-03-04. openai.com ↗

- 34 Infobae. *Programa piloto con IA y tecnología lleva a 171 escuelas salvadoreñas a niveles de países desarrollados en PISA, dice el gobierno*. 2026-09-09. [infobae.com](https://www.infobae.com) ↗
- 35 La Noticia SV. *PISA 2025: El Salvador mejora sus resultados en Ciencias, Matemáticas y Lectura*. 2026-09. lanoticiasv.com ↗
- 36 Allcott, H., Baron, E. J., Dee, T., Duckworth, A. L., Gentzkow, M. y Jacob, B.. *The Effects of School Phone Bans: National Evidence from Lockable Pouches*. 2026-04. nber.org ↗
- 37 Rest of World. *South Korea's AI textbooks fail after rushed rollout*. 2025. restof-world.org ↗
- 38 CNBC. *Key AI hub China restricts schoolchildren's use of the tech*. 2025-05-15. cnbc.com ↗
- 39 Lichand, G., Moreno-Louzada, L., da Costa, T. y Gentzkow, M.. *The Educational Impacts of School Phone Bans: Evidence from Brazil*. 2026-05. nber.org ↗
- 40 Figlio, D. N. y Özek, U.. *The Impact of Cellphone Bans in Schools on Student Outcomes: Evidence from Florida*. 2025-10. nber.org ↗
- 41 Instituto Nacional de Evaluación Educativa. *Resultados de TALIS 2024: informe español*. 2025-10-07. inee.educacion.es ↗
- 42 Ellington, A. J.. *A Meta-Analysis of the Effects of Calculators on Students' Achievement and Attitude Levels in Precollege Mathematics Classes*. 2003. psycnet.apa.org ↗
- 43 Gortazar, L. y Montalbán, J.. *La nueva ley de ratios: estimación de su impacto económico*. 2026-05-26. esade.edu ↗
- 44 Roediger, H. L. y Karpicke, J. D.. *Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention*. 2006. doi.org ↗