

Evaluación del efecto de la inteligencia artificial en el aprendizaje

INFORME COMPLETO

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

Índice general

Antes de empezar	1
1 El examen sin la máquina	2
1.1. Treinta meses en un condado chino	2
1.2. Cuatro de cada cinco	3
1.3. El mismo resultado, medido por dentro	4
1.4. Lo que dicen los ensayos con sorteo	6
1.5. Se rompe el termómetro	6
1.6. Por qué el alumno no lo nota	7
1.7. Por qué estos veredictos	8
2 PISA 2025 de cerca	9
2.1. El resultado	9
2.2. Dónde está el problema español	11
2.3. Lo que la OCDE dice que ha pasado	12
2.4. Dos ideas cómodas que el informe no respalda	13
2.5. Qué pinta aquí la IA	14
2.6. Quién aguanta	15
2.7. Por qué estos veredictos	16
3 Peor que hace quince años, no que hace cien	17
3.1. Un siglo	17
3.2. Quince años	18
3.3. Quién se examina	19
3.4. Fuera de los países ricos	19
4 El alumno que no quiere	21
4.1. Dos cursos con el tutor a un clic	21
4.2. El patrón es anterior a la IA	22
4.3. Lo que dicen de sí mismos y lo que hacen	23
4.4. Lo que compran las familias	24

4.5.	El tutor que no espera a que le pregunten	25
4.6.	Por qué estos veredictos	26
5	Los ensayos, uno por uno	27
5.1.	Un artículo retirado	27
5.2.	La vara de medir	29
5.3.	Qué se ha ensayado	29
5.4.	La regularidad: la mejora vive en la medida cercana	31
5.5.	Donde sí ha funcionado	32
5.6.	Cinco preguntas para leer el próximo titular	32
5.7.	Por qué estos veredictos	33
6	Milagros y marchas atrás	34
6.1.	El Salvador	34
6.2.	Corea del Sur	35
6.3.	Estonia	36
6.4.	China	36
6.5.	La bala de plata anterior	37
6.6.	Por qué estos veredictos	37
7	El profesor	39
7.1.	Quién usa qué	39
7.2.	Detrás del adulto, no delante del alumno	40
7.3.	Lo que cambia es la evaluación	41
7.4.	Por qué estos veredictos	42
8	España	43
8.1.	El retrato	43
8.2.	La apuesta: bajar la ratio	44
8.3.	Lo que hacen las familias	45
8.4.	Donde la IA tendría sentido	45
8.5.	Las comunidades	45
8.6.	Los que llegan ahora	46
8.7.	Por qué estos veredictos	46
9	Cómo se hace aprender a quien no quiere	47
9.1.	Lo que dicen juntos los capítulos anteriores	47
9.2.	Uno: la obligación la pone una persona	48

9.3. Dos: el examen sin máquina, corto y frecuente	48
9.4. Tres: mirar el tiempo, no la nota	49
9.5. Cuatro: primero sin, luego con	49
9.6. Cinco: si hay tutor de IA, que no espere a que le llamen	49
9.7. Seis: la IA, detrás del adulto	50
9.8. Lo que no va a pasar	51
9.9. Por qué estos veredictos	51
10 Qué hacer	52
10.1. Tres conclusiones con las que trabajar unos años	52
10.2. Si estudia: bachillerato, universidad, una oposición	53
10.3. Si tiene hijos en edad escolar	54
10.4. Si enseña	54
10.5. Si decide sobre un centro o sobre un presupuesto	55
10.6. Lo que no sirve	55
10.7. Lo que queda por encima de todo esto	55
Síntesis de veredictos	57
Referencias	61

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.

Conviene saber quién habla. De educación opina todo el mundo, y casi siempre en tono de derrumbe: cada generación ha creído que la siguiente se echaba a perder. A veces era verdad y casi nunca del todo. Quien vende la herramienta promete una revolución; quien la teme cita el peor estudio. Los ensayos bien hechos son pocos, recientes y más modestos que los titulares en ambos sentidos.

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, con el dato más reciente que existe. 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.

El examen sin la máquina

EN BREVE

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 el informe, empezando por su dato más duro. Usar IA para los deberes sube la nota de los deberes y baja la del examen: se ve a los seis meses en el examen mensual, y en los exámenes de acceso sigue creciendo unos dos años. No es que la herramienta atonte: es que cuatro de cada cinco alumnos la usan para acabar antes, y acabar antes era justo lo que no enseñaba. El primer daño no es al aprendizaje, sino al termómetro: la nota de lo hecho en casa ha dejado de decir lo que el alumno sabe.

—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

1.1 Treinta meses en un condado chino

El mejor dato que existe sobre lo que hace la IA con un alumno corriente no sale de un laboratorio. Strömberg, Lei y Wu, tres economistas de Estocolmo y Hong Kong, obtuvieron treinta meses de registros, de sep-

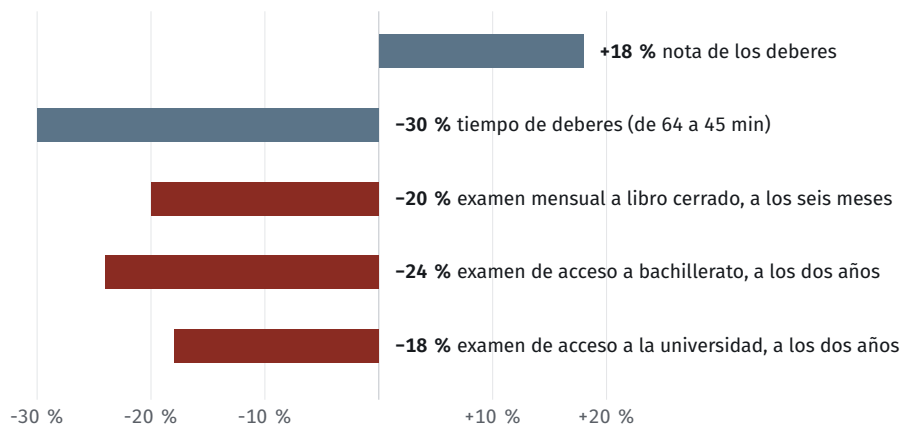
tiembre de 2022 a junio de 2025, de los 26.811 alumnos de 7.º a 12.º de un condado del centro de China: nota y duración de los deberes de cada semana, exámenes mensuales a libro cerrado en nueve asignaturas, y los dos exámenes que allí deciden una vida, el de acceso a bachillerato y el de acceso a la universidad. Los profesores pidieron a los alumnos que miraran en sus aplicaciones la fecha en que se registraron. Al empezar el periodo casi nadie usaba IA generativa; al acabar, el 80 %. Como cada uno empezó en un momento distinto, se puede comparar cómo le cambia la vida escolar a quien empieza con cómo le va, esos mismos meses, a quien todavía no ha empezado. Antes de adoptar la herramienta, los dos grupos tenían notas casi idénticas. No hay sorteo: el resultado vale lo que valga el supuesto de que, sin la herramienta, unos y otros habrían seguido igual.¹

A los seis meses, los deberes pasan de 64 a 45 minutos y su nota sube un 18 %. En el examen mensual, sin la herramienta, el efecto es pequeño al principio y se completa hacia los seis meses: un 20 % menos respecto a la media anterior. En los exámenes de acceso, que recogen lo aprendido en varios cursos, la pérdida tarda unos dos años en completarse: un 24 % en el de bachillerato y un 18 % en el de la universidad. Los autores sacan de ahí una advertencia que afecta a casi todo lo que se ha publicado: los estudios de pocas semanas «pueden subestimar sistemáticamente» el coste.

Las herramientas eran las de cualquiera: Doubao y DeepSeek, asistentes generales sin ninguna función educativa. Y pierden más quienes menos se esperaba. Las ciencias sociales caen un 27 %, las ciencias y las matemáticas un 22 %, el inglés un 17 %, la lengua un 9 %. Caen más los pequeños que los mayores, los chicos que las chicas y, sobre todo, los que antes iban mejor. En el trabajo, la IA iguala porque sube a los de abajo. En la escuela iguala porque baja a los de arriba.

1.2 Cuatro de cada cinco

El estudio separa dos maneras de usarla, y ahí está lo importante. Cuatro de cada cinco usuarios acaban los deberes en menos de cincuenta minutos —más deprisa que el más rápido de quienes no usan IA—, con notas que

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.

coinciden con lo que sabe hacer un chatbot, y con exámenes claramente peores. El otro usuario de cada cinco sigue dedicando a los deberes el mismo tiempo que antes, saca mejor nota en ellos y exactamente la misma que los demás en el examen. Con las palabras de los autores: cuando los alumnos con IA dedican a los deberes el mismo tiempo que los alumnos sin IA, aprenden lo mismo.

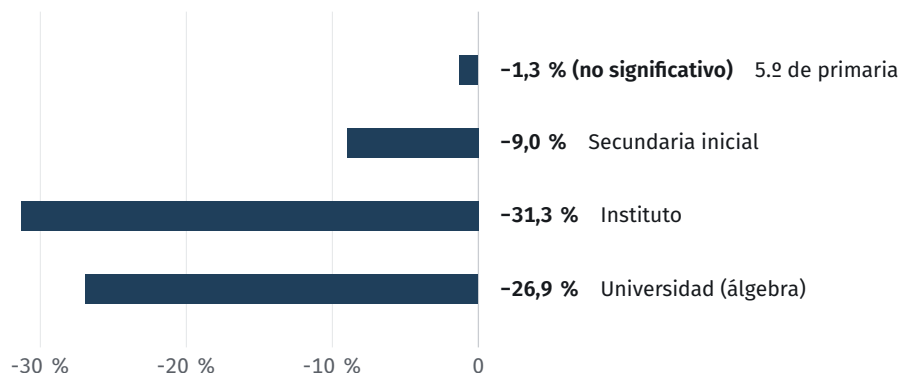
La herramienta, entonces, no parece dañar por sí misma. Daña el uso que elige la inmensa mayoría cuando nadie mira, con una cautela que los propios autores señalan: quien decide mantener el tiempo de trabajo puede ser, ya de entrada, otro tipo de alumno. Los autores lo dicen sin adornos: existen tutores bien diseñados, muchos gratuitos, y «la mayoría de los alumnos sencillamente no los usa»; prefiere la herramienta general, que hace fácil quitarse los deberes de encima. Y añaden que el problema «no desaparecerá con los avances de la técnica», porque no es técnico.

1.3 El mismo resultado, medido por dentro

ALEKS, una plataforma estadounidense de matemáticas que usan más de cuatro millones de alumnos al año, guarda cuánto tarda cada uno en cada problema desde 2015. Sus problemas son de dos clases: de enunciado, que

se copian y se pegan en un chatbot, y de gráficas, que hay que manipular en pantalla. Desde finales de 2022, el tiempo dedicado a los primeros cae trimestre a trimestre respecto a los segundos.²

Figura 2 Caída del tiempo por problema en los ejercicios que se pueden pegar en un chatbot



Rismanchian et al. (2026), 3,2 millones de interacciones en la plataforma ALEKS, 2015-2025. Variación acumulada en once trimestres respecto a los problemas de gráficas. Preprint sin revisión por pares.

Tres detalles convierten la curva en una prueba. El efecto crece con la edad: nada en 5.º de primaria, un 9 % en los primeros cursos de secundaria, un 31 % en el instituto. En las pruebas de nivel hechas sin vigilancia, los alumnos también van más deprisa; con un vigilante delante, el tiempo vuelve a ser el de siempre. Y en esas pruebas vigiladas, la razón de probabilidades de acertar los problemas de enunciado ha bajado un 25 % en tres años —para quien partiera de un 80 % de aciertos, equivale a quedarse en un 75 %—, mientras que en las no vigiladas ha subido. La significación estadística de ese último dato es justa y conviene no exagerarlo; la dirección es la del estudio chino. Los autores distinguen dos cosas que se confunden a menudo. Delegar en una calculadora la aritmética y conservar el planteamiento es descargar trabajo. Pegar el enunciado y copiar la respuesta es otra cosa, que llaman rendición: el alumno se salta el esfuerzo de recuperar lo que sabe, que es lo que convierte haber visto algo en recordarlo.

1.4 Lo que dicen los ensayos con sorteo

En un instituto turco, el equipo de Hamsa Bastani, de la Universidad de Pensilvania, puso a cerca de mil alumnos a practicar matemáticas en tres condiciones asignadas al azar. Con un asistente sin restricciones, los alumnos resolvieron un 48 % más de ejercicios durante la práctica y, en el examen posterior sin ayuda, rindieron un 17 % peor que quienes habían practicado solos. Con un asistente preparado para dar pistas y no soluciones resolvieron un 127 % más durante la práctica, y en el examen no perdieron nada. Tampoco ganaron: el tutor bien hecho evitó el daño y no produjo mejora.³

La síntesis de la OCDE de enero de 2026 resume así el conjunto de la evidencia: quienes usan herramientas generales entregan trabajos mejores, y esa ventaja «desaparece, y a veces se invierte, en los exámenes, cuando se retira el acceso». Completar una tarea con IA, dice, no lleva por sí mismo a aprender.⁴

1.5 Se rompe el termómetro

Entre los alumnos chinos que no usan IA, quien saca mejor nota en los deberes saca mejor nota en el examen, como ha ocurrido siempre. Entre los que la usan, la relación se invierte: a mejor nota de deberes, peor examen. Una subida rápida de las notas de casa ha pasado a ser una señal de alarma.¹

En una universidad israelí, Hausman, Rigbi y Weisburd siguieron a 36.000 alumnos de 2018 a 2024, y ocurre lo mismo con otra forma. En las asignaturas que se evalúan con trabajos hechos en casa, las notas suben, sobre todo las de los alumnos flojos, y los suspensos caen un tercio. En las que se evalúan con examen presencial no cambia nada. Antes de ChatGPT, el puesto de un alumno en unas predecía su puesto en las otras; después, cada vez menos. Y quienes tuvieron IA desde primero rinden luego mejor en las asignaturas de trabajos y no mejor, a veces peor, en las de examen de los cursos superiores.⁵

El fenómeno tiene nombre desde 1979. El psicólogo Donald Campbell advirtió de que, cuanto más se usa un indicador para decidir, más se

corrompe: quien es juzgado por una cifra acaba produciendo la cifra.⁶ La nota de los deberes servía porque hacerlos costaba lo mismo que aprender. Desde que una máquina los hace, mide a la máquina. Hipótesis de este informe: el primer daño de la IA en la escuela no es al aprendizaje, sino a la medida. Profesores y padres miran la nota de los deberes y de los trabajos para saber cómo va un alumno, y esa nota ha dejado de informar justo cuando más falta hacía. Hasta que llega un examen sin máquina, todo parece ir mejor que nunca; y donde el único examen serio es el de final de etapa, la pérdida puede tardar dos años en verse entera.

1.6 Por qué el alumno no lo nota

En 2019, Louis Deslauriers y sus colegas de Harvard repartieron al azar a los alumnos de Física entre una clase magistral pulida y una sesión de trabajo activo, con el mismo material. Los de la clase magistral creyeron haber aprendido más. Los de la sesión activa, que la encontraron más incómoda, sacaron mejor nota. La sensación de aprender y el aprendizaje iban en sentidos opuestos.⁷ El psicólogo Robert Bjork lleva treinta años documentándolo: lo que hace más fluida la práctica —releer, tener la solución a la vista— mejora el rendimiento del momento y empeora lo que queda, y las «dificultades deseables», como esforzarse en recordar, hacen lo contrario.⁸ Un chatbot es la explicación más fluida que se ha inventado: leer su solución produce la sensación exacta de haberla entendido. Nada en la experiencia diaria del alumno le avisa de lo que pasará en el examen.

Hay un contrapeso, y afina el mecanismo. En varios ensayos de un equipo en el que figura la psicóloga Angela Duckworth, con adultos que aprendían a redactar cartas de presentación, quienes practicaron con una herramienta de IA se esforzaron menos y escribieron mejor después, sin ella. Bastaba ver el texto corregido por la máquina para mejorar lo mismo: la IA enseña con el ejemplo.⁹ No es el sudor lo que enseña. Es ver un buen modelo y tener que producir después sin ayuda. El alumno que pega la solución también ha visto el modelo. No lo mira, y nadie le pide después que lo haga solo.

1.7 Por qué estos veredictos

Los dos primeros son «muy probables» porque ya se miden, en dos países y con métodos distintos, y porque su causa no es técnica: mientras acabar antes sea posible y nadie compruebe después, la mayoría acabará antes. La caída en exámenes nacionales se queda en «probable»: el mecanismo está medido en un condado chino y en una plataforma, no en un país, y los exámenes nacionales recogen muchas más cosas. La generación «incapaz de pensar» es «muy improbable» por lo que enseña el quinto alumno: quien mantiene el tiempo de trabajo aprende lo mismo que antes. 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
Caída apreciable en exámenes nacionales sin IA atribuible a su uso en los deberes	Probable Grave y nacional; recuperable
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.

— Si esto es así, las notas de un país deberían notarlo. El examen que acaba de publicarse lo permite comprobar, y dice algo más complicado que los titulares.

PISA 2025 de cerca

EN BREVE

Es el peor resultado desde que existe la prueba, en España y en casi todas partes, y se ha achacado en parte a la IA. La OCDE no lo hace: en su explicación de la caída no la nombra ni una vez. La lectura cae desde 2012 y las matemáticas desde 2018. Lo que el examen retrata es una pérdida de atención sostenida y de ganas, anterior a ChatGPT. La IA llega a un alumno que ya leía menos. PISA no permite medir cuánto añade; la pista está en un sitio concreto, los deberes.

—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^{10, 11}

0

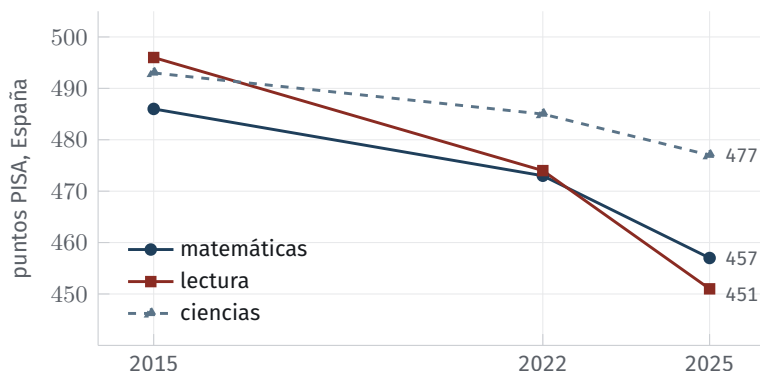
veces que aparecen «IA», «inteligencia artificial» o «ChatGPT» en el recuadro con el que la OCDE explica por qué ha caído la lectura¹²

—11 %

tiempo diario de deberes que declaran los alumnos de la OCDE respecto a 2022; baja en casi todos los países a la vez¹³

2.1 El resultado

PISA 2025 examinó a 760.000 alumnos de 15 años de 91 países y economías. La media de la OCDE es la más baja de la serie en las tres materias: 482 en ciencias, 461 en lectura y 463 en matemáticas, tres, catorce y nueve puntos menos que en 2022. Uno de cada cinco alumnos no alcanza el nivel básico en ninguna de las tres.¹⁰ Como referencia: la OCDE equipara los 28 puntos perdidos en lectura en una década a un curso y medio, de modo que un curso viene a ser unos veinte puntos.¹⁴

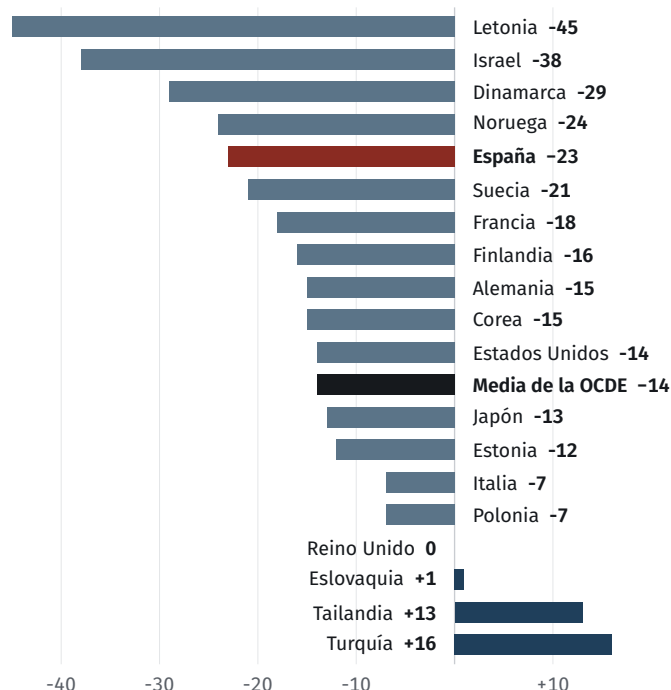
Figura 3 España en PISA, 2015-2025

OCDE, PISA 2015, 2022 y 2025. La caída de ciencias entre 2022 y 2025 no es estadísticamente significativa.

España obtiene 477 en ciencias, 451 en lectura y 457 en matemáticas. Respecto a 2022 pierde 23 puntos en lectura y 16 en matemáticas; en ciencias, la OCDE considera el resultado «más o menos igual». Respecto a 2015, la lectura ha caído 45 puntos: más de dos cursos.^{11,15} La muestra fueron unos 30.000 alumnos de 956 centros; el 77 % cursaba 4.º de ESO, de modo que casi uno de cada cuatro iba retrasado de curso. El Ministerio tituló su nota de prensa con que el alumnado español declara más perseverancia que la media de la OCDE.

Hay una advertencia técnica que casi nadie ha recogido. Las exclusiones de alumnos subieron en España al 7,9 %, sobre todo por Cataluña, donde pasaron del 5,6 al 23,2 %, y Murcia, del 4,1 al 12,7 %. Los resultados catalanes no se publican por separado, y la OCDE avisa de que un nivel así de exclusiones sesga «muy probablemente» el resultado al alza. El dato español real puede ser algo peor que el publicado.¹⁵

Por comunidades, según los datos difundidos por la prensa, Madrid, Castilla y León y Asturias quedan arriba, con 495, 493 y 492 puntos en ciencias, y Melilla y Ceuta abajo, a 90 y 75 puntos de Madrid. La Comunidad Valenciana pierde 58 puntos en lectura y el País Vasco 47. Andalucía pierde 16 en matemáticas, 20 en lectura, donde queda en 441, y 11 en ciencias.¹⁶

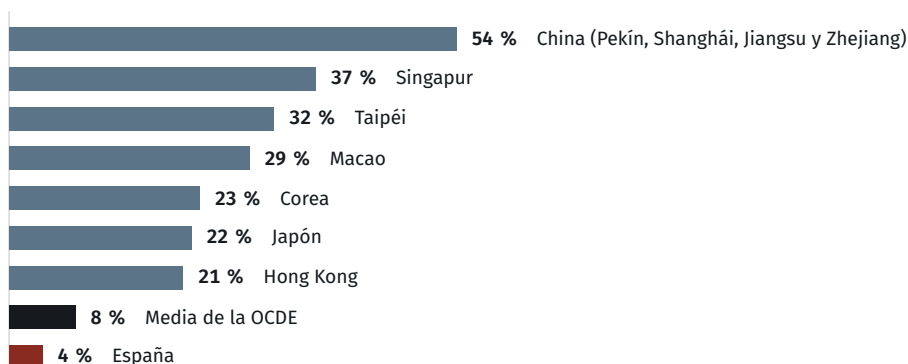
Figura 4 Variación de la puntuación en lectura entre 2022 y 2025

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

2.2 Dónde está el problema español

España no tiene más alumnos hundidos que los demás. Alcanzan el nivel básico el 76 % en ciencias, el 67 % en matemáticas y el 68 % en lectura, frente al 74, 65 y 69 % de la OCDE. Lo que tiene es la mitad de alumnos excelentes: un 4 % en ciencias, un 4 % en matemáticas y un 2 % en lectura, frente al 7, 8 y 6 %. En las cuatro provincias chinas que participan, el 54 % de los alumnos alcanza la excelencia en matemáticas; en Singapur, el 37 %. El 44 % de los alumnos españoles pertenece al cuarto más favorecido de todos los examinados en el mundo, y con ese origen obtiene 509 puntos en ciencias; con el mismo origen, en Singapur, Taipéi o Macao se saca bastante más.¹⁵

La desigualdad, en cambio, es menor que la media: 71 puntos entre el cuarto más favorecido y el menos favorecido, frente a 85. Pero se ha reducido por el lado malo. Entre 2015 y 2025, dice la OCDE, el rendimiento

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

OCDE, PISA 2025, nota de país de España.

bajó más entre los alumnos favorecidos que entre los desfavorecidos, en España y en el conjunto. Es el mismo dibujo que en el condado chino: la caída reciente comprime desde arriba.

2.3 Lo que la OCDE dice que ha pasado

El informe dedica un recuadro largo a una pregunta: por qué cae la lectura. Distingue cuatro fuerzas, y ninguna es la IA.¹²

El alumno aguanta menos. No se hunden todas las destrezas: las ciencias resisten mientras la lectura cae. Lo que más baja son las tareas con textos largos y las que piden evaluar, relacionar varias fuentes y reflexionar; lo que menos, localizar un dato suelto. En los ejercicios de fluidez del principio de la prueba, los lectores «apresurados» —responden rápido y mal— pasan del 7 % en 2018 al 11 % en 2025 en la media de la OCDE. Hay más respuestas rápidas e incorrectas y menos preguntas en blanco: se contesta cualquier cosa. El rendimiento cae más en el tramo final de la prueba que en el inicial. Y los alumnos declaran haberse esforzado menos en ella que en 2022 y en 2018. Entre países, el cambio en la proporción de alumnos que se dicen curiosos se correlaciona con el cambio en lectura ($r = 0,54$).

El aula es más difícil. Según la encuesta de la OCDE al profesorado, la parte de la clase que se va en mantener el orden ha subido del 13 %

en 2018 al 16 % en 2024. Menos alumnos dicen que su profesor sigue explicando hasta que todos entienden. Los directores hablan de más insultos, intimidación y vandalismo.

El sistema responde tarde. Tres de cada cuatro profesores enseñan a grupos muy heterogéneos, con más alumnos recién llegados, más necesidades especiales y menos apoyos. La OCDE precisa que el cambio en la composición del alumnado «no explica la caída general».

La sociedad lee menos. El gusto por la lectura baja desde 2009, y también entre los padres: leer ha perdido sitio como afición de los adultos. Leer ficción, textos largos y en papel se asocia a leer mejor. Los examinados en 2025 tenían diez u once años en 2020, la edad en que se pasa de aprender a leer a leer para aprender; pero Suecia no cerró las escuelas y pierde 21 puntos, y Costa Rica las cerró veinte semanas y no pierde nada. La OCDE describe la pandemia como un catalizador de cambios de hábito, no como su causa.

La conclusión es literal: la evidencia «no apunta a una explicación sencilla», sino a fuerzas pequeñas que se alinean.

2.4 Dos ideas cómodas que el informe no respalda

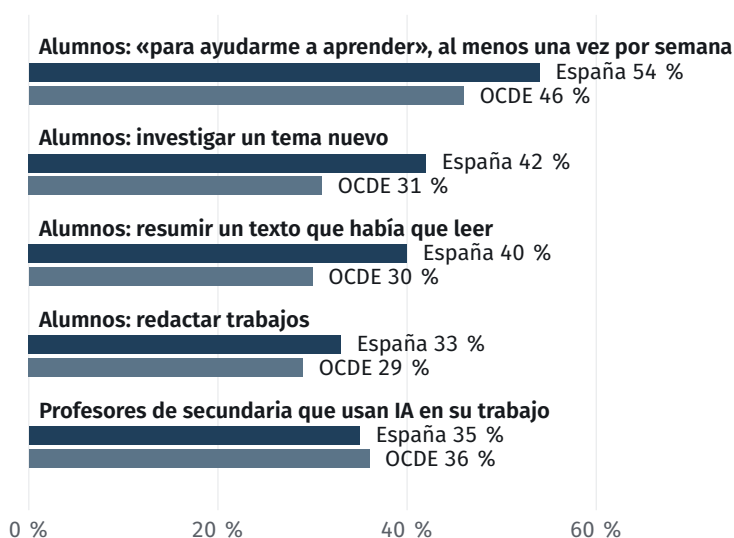
«Es que cada vez pasan más horas con pantallas.» Entre 2022 y 2025, el tiempo que los alumnos declaran pasar con dispositivos fuera de la escuela bajó, para estudiar y para el ocio. Los que pasan más de dos horas diarias de ocio digital son seis puntos menos; los de más de siete horas en fin de semana, 3,6 puntos menos. Sube, eso sí, la parte dedicada a redes sociales. Las notas cayeron mientras caía el tiempo de pantalla.¹³

«Hay que prohibir el móvil.» España ya lo hizo. El 86 % de sus alumnos va a centros donde el móvil está prohibido en el recinto; en 2022 era el 67 %, y la media de la OCDE es el 49 %.¹⁵ En esos tres años perdió 23 puntos en lectura. Eso no dice qué habría pasado sin la prohibición; dice que no bastó. La OCDE encuentra que los centros con prohibición tienen menos distracción, y que su relación con el rendimiento es «poco clara»: sus alumnos puntúan igual o peor. El capítulo 6 recoge los estudios causales: entre nada y una mejora pequeña, según el país y los años que lleve la medida.

2.5 Qué pinta aquí la IA

PISA preguntó por primera vez por los chatbots. En la media de la OCDE solo el 14 % de los alumnos no los ha usado nunca para tareas escolares, y uno de cada cinco los usa casi a diario. España está por encima: el 54 % los usa al menos una vez por semana «para ayudarse a aprender», frente al 46 % de la OCDE; el 42 % para investigar un tema nuevo (31 %); el 40 % para resumir un texto que había que leer (30 %); el 33 % para redactar trabajos (29 %). Solo el 8 % dice no usarlos nunca.^{13, 15}

Figura 6 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).

Quienes no usan IA para tareas concretas —resumir, redactar, buscar— puntúan más que quienes la usan. Para el uso genérico «para ayudarme a aprender», los usuarios semanales rinden igual que los no usuarios, y los que la usan casi a diario, peor. La OCDE insiste en que es una asociación, que puede reflejar tanto quién recurre a la herramienta como lo que la herramienta hace. Añade un matiz que orienta: seis de cada diez alumnos han tenido que evaluar en clase la calidad de algo generado por una IA, y entre los usuarios frecuentes, quienes han recibido esa enseñanza

rinden algo mejor. Esa enseñanza es más común entre los favorecidos. Conclusión del informe: ampliar el acceso, sin más, «es improbable que produzca beneficios».¹³

Y hay un dato que casi nadie ha citado. Los alumnos de la OCDE declaran 1,4 horas diarias de deberes, unos diez minutos menos que en 2022: un 11 % menos en tres años, «en casi todos los sistemas educativos» a la vez. La OCDE no lo relaciona con nada. Este informe propone una relación que PISA no puede probar: son los tres años en que llegó ChatGPT, y el estudio chino mide por dentro exactamente eso, deberes que pasan de 64 a 45 minutos.

Lo que los datos permiten afirmar es poco y es firme: la caída empezó antes. La lectura baja desde 2012 y las matemáticas desde 2018, y el examen de 2025 retrata a alumnos que llevaban dos cursos y medio con chatbot, no diez. Cuánto ha añadido la IA entre 2022 y 2025, PISA no lo puede decir: no es un experimento, y que la OCDE no la nombre tampoco prueba que no pese. Hipótesis de este informe: la IA no originó la caída; la encontró hecha y actúa como acelerante sobre la pieza que ya fallaba, el esfuerzo sostenido, allí donde se la deja sola, que es en casa. Si es así, el próximo examen, que será el de 2029, saldrá peor que este aunque no cambie nada más, porque esos alumnos habrán hecho con IA toda la secundaria.

2.6 Quién aguanta

El Reino Unido no pierde nada en lectura y gana doce puntos en ciencias. Turquía sube dieciséis en lectura; Eslovaquia, uno. Estonia, el primer país europeo, pierde doce. Los nórdicos, los sistemas más digitalizados de Europa, se hunden: Dinamarca pierde 29 puntos en lectura, Noruega 24, Suecia 21, Finlandia 16. La OCDE recuerda que la riqueza explica solo la mitad de las diferencias entre países: Luxemburgo gasta 288.521 dólares por alumno y rinde menos que Irlanda, Francia o Italia, que gastan menos de la mitad.¹⁰ Nadie ha demostrado por qué aguanta. Lo que Inglaterra ha hecho desde 2010 es conocido —enseñanza sistemática de la lectura, un currículo de contenidos, exámenes externos—, y que esa sea la causa es una conjetura razonable, no un resultado.

2.7 Por qué estos veredictos

Que PISA 2029 salga igual o peor es «probable» por dos razones que se suman: las fuerzas que la OCDE describe —menos lectura, menos aguante, aulas más difíciles— no han cambiado de signo, y quienes se examinen habrán hecho toda la secundaria con un chatbot. No sube a «muy probable» porque tres años bastan para que un país cambie la evaluación, y el Reino Unido demuestra que se puede no caer. La IA como causa principal de la caída de 2025 es «muy improbable» por pura cronología: la lectura baja desde 2012. La recuperación de algún país grande es solo «posible»: nadie ha demostrado todavía qué la produce.

VEREDICTO Hasta 2031	
PISA 2029 con resultados iguales o peores que los de 2025 en la media de la OCDE	Probable Grave; sería la cuarta caída seguida
España sigue por debajo de la media de la OCDE en las tres materias en PISA 2029	Probable Grave y nacional
La IA, y no otra cosa, como causa principal de la caída medida en 2025	Muy improbable La caída empieza diez años antes
Recuperación apreciable en algún país grande que cambie evaluación y deberes	Posible El Reino Unido indica que se puede no caer

Peor que hace quince años, no que hace cien

EN BREVE

«Mínimo histórico» significa mínimo desde el año 2000, que es cuando empezó PISA. Medido en un siglo, nunca ha sabido leer tanta gente ni ha estado tanta gente tantos años en la escuela. Medido en quince años, se ha perdido una parte de lo ganado, y la pérdida empezó antes del móvil en unos países y antes de la IA en todos. Las dos cosas son ciertas a la vez, y fuera de los países ricos el problema es otro: que no hay profesor.

48 %

españoles mayores de diez años que sabían leer y escribir en 1910; en 1970 eran el 91 %¹⁷

256

puntuación en lectura de los estadounidenses de 13 años en 2023: un punto más que en 1971, siete menos que en el máximo de 2012¹⁸

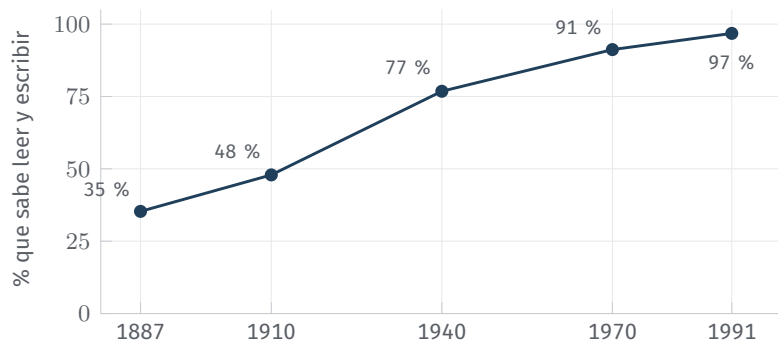
272

millones

niños y jóvenes sin escolarizar en el mundo¹⁹

3.1 Un siglo

En el censo de 1887 sabía leer y escribir el 35 % de los españoles mayores de diez años. En 1910, el 48 %: más de la mitad del país no podía leer una carta. En 1940 eran el 77 %, en 1970 el 91 % y en 1991 el 97 %.¹⁷ La escolaridad obligatoria hasta los 14 años se aprobó en 1964 y se generalizó con la EGB, a partir de 1970; hasta los 16, es de 1990. La escuela para todos hasta la adolescencia es, en España, un invento de hace una generación.

Figura 7 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.

Durante el siglo XX, además, las puntuaciones en las pruebas de inteligencia subieron en los países industrializados unos tres puntos por década: es el efecto que lleva el nombre de James Flynn, el investigador que lo documentó. Nadie cree que los bisabuelos fueran torpes: cambiaron la escuela, la alimentación y el tipo de problemas que la vida pedía resolver.

3.2 Quince años

Esa subida se ha parado, y en algunos sitios se ha dado la vuelta. Los registros de más de 730.000 reclutas noruegos muestran que las puntuaciones alcanzan el máximo entre los nacidos hacia 1975 y bajan desde entonces. La bajada aparece dentro de las familias —el hermano menor puntúa menos que el mayor—, de modo que no se debe a quién tiene hijos ni a la inmigración, sino al ambiente.²⁰ Los nacidos en 1975 cumplieron quince años en 1990: mucho antes del teléfono móvil.

Estados Unidos examina a sus alumnos de 13 años con la misma prueba desde 1971. En lectura sacaron 255 puntos aquel año, 263 en 2012 y 256 en 2023. En matemáticas, 266 en 1973, 285 en 2012 y 271 en 2023. Cuarenta años de mejora lenta y once de caída, que empieza ocho años antes de la pandemia y diez antes de ChatGPT. Quienes más pierden son los de abajo, y los que leen por gusto casi a diario han pasado del 27 al 14 %.¹⁸ La edición de 2025 no cambia el cuadro a esa edad: a los 13 años no hay

variación significativa respecto a 2023, y a los 9, lectura y matemáticas suben cuatro puntos respecto a 2022.²¹

PISA cuenta lo mismo con menos perspectiva: la lectura baja desde su máximo de 2012 y las matemáticas desde 2018. «El peor resultado de la historia» es el peor de veinticinco años.

3.3 Quién se examina

Hay un matiz que da la razón a quien sospecha de las comparaciones. Cuando un país escolariza a quienes antes dejaba fuera, sus medias bajan aunque enseñe mejor, porque entran en el examen alumnos que antes no contaban. En Camboya, Ecuador, El Salvador, Líbano, Marruecos y Paraguay, la parte de los jóvenes de 15 años que PISA puede examinar ha subido entre 11 y 24 puntos en una década. De los países en esa situación, solo tres mejoran además en alguna materia: Camboya, la República Dominicana y El Salvador.¹² En los países ricos, donde casi todos estaban ya dentro, ese efecto no explica la caída, y la OCDE lo comprueba.

3.4 Fuera de los países ricos

Todo lo que discute este informe —si el chatbot le hace los deberes al alumno— es una preocupación de quien ya tiene escuela. Hay 272 millones de niños y jóvenes sin escolarizar.¹⁹ En los países de renta baja y media, siete de cada diez niños de diez años no entienden un texto sencillo.²² Para escolarizar a todos en primaria y secundaria harán falta, de aquí a 2030, 44 millones de profesores más, contando los que hay que reponer; quince de ellos en el África subsahariana, donde un tercio de las escuelas rurales no tiene electricidad fiable y dos tercios no tienen internet.^{23, 24}

Allí el signo de la IA cambia. En Tennessee, un tutor automático compete con un profesor titulado y una clase de refuerzo, y añade muy poco. En Benín City, en Nigeria, compete con no tener a nadie por las tardes, y en seis semanas logra, según sus autores, el equivalente a casi dos cursos de aprendizaje ordinario, como se verá en el capítulo 5.²⁵ La misma he-

ramienta que en un país rico sustituye esfuerzo, en uno pobre sustituye ausencia.

LO QUE SE SIGUE

No estamos ante la generación más ignorante de la historia. En varias pruebas y en varios países, los alumnos de hoy rinden peor que los de hace quince años, después de un siglo de avances enormes en escolarización. Es grave porque interrumpe una mejora muy larga, y es recuperable por lo mismo: lo que se ganó con escuela, libros y tiempo se puede volver a ganar. La pregunta útil no es si la IA acaba con la educación, sino si ayuda a recuperar lo perdido o lo agranda. De eso tratan los capítulos siguientes.

El alumno que no quiere

EN BREVE

El tutor que no da la respuesta existe, funciona y está a un clic. Casi nadie lo usa. En los cuatro programas de los que hay datos, cada uno medido a su manera, el uso real es bajo o superficial: según la medida, entre una vigésima parte y un tercio. Ninguna de esas medidas dice directamente cuántos alumnos se esfuerzan. La tesis de este informe es que lo escaso no es la explicación, que hoy es gratis, sino un motivo para sentarse. Casa con un dato español: casi la mitad de las familias sigue pagando clases particulares.

17 %

sesiones con errores en las que el alumno mediano escribió algo a Khanmigo, casi siempre una respuesta suelta o un clic, en un ensayo de dos cursos en 18 institutos de Tennessee; el 96 % lo había probado alguna vez²⁶

60

sistemas educativos donde ha subido desde 2022 la parte de alumnos a quienes aprender cosas nuevas les aburre; no ha bajado en ninguno¹⁰

47 %

hogares españoles con hijos en edad escolar que pagan clases particulares: 2.782 millones de euros al año²⁷

4.1 Dos cursos con el tutor a un clic

El ensayo más serio sobre un tutor de IA en una escuela corriente se hizo en Tennessee. Alumnos de bajo rendimiento de 18 institutos fueron asignados al azar, durante dos cursos y en horario lectivo, a practicar matemáticas con Khan Academy y su tutor Khanmigo, configurado para

orientar y no dar la respuesta. El grupo de comparación seguía con otros programas y con refuerzo en grupo pequeño.²⁶

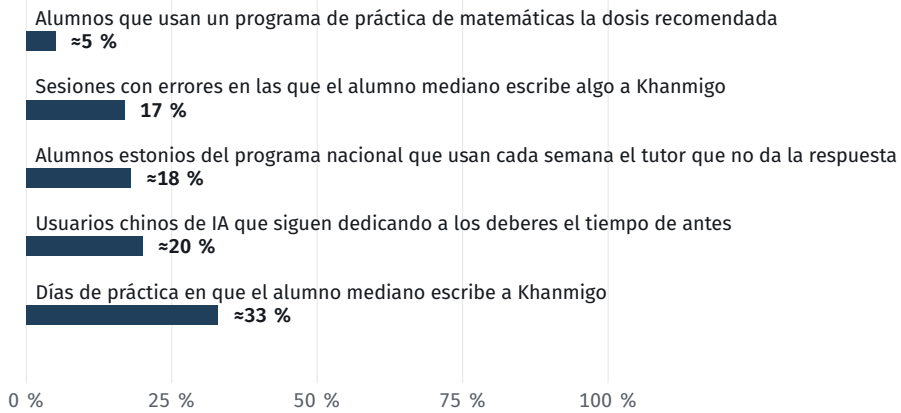
El programa sirvió: sus alumnos ganaron 1,3 puntos de percentil nacional por trimestre, entre 0,06 y 0,08 desviaciones típicas al año. Pero es difícil atribuir esa ganancia al tutor, porque apenas se usó. El 96 % lo probó alguna vez. El alumno mediano le escribió uno de cada tres días de práctica, y solo en el 17 % de las sesiones en las que cometía errores; la mayoría de esos mensajes eran una respuesta suelta o un clic en una sugerencia. El ensayo no separa el efecto del tutor del de la plataforma: compara el paquete entero con lo que había. La conclusión de los autores cabe en una línea: hacer realidad la promesa de la tutoría con IA exige conseguir que los alumnos la usen, no darles acceso.

Philip Oreopoulos, que firma el estudio, lo explicó a la prensa especializada: si un alumno con ganas quiere usar la herramienta como un buen tutor, está lista; el problema es que la mayoría no la usa así, y la misma técnica que sirve de tutor sirve para hacerse la vida más fácil, que es menos esfuerzo y menos aprendizaje. Sal Khan, fundador de la plataforma, reconoció que para la mayoría el tutor había sido un «no acontecimiento» y que el ensayo ponía el foco en cuánto importa la motivación.²⁸

4.2 El patrón es anterior a la IA

En 2024, un análisis de los programas de práctica de matemáticas en línea más usados en Estados Unidos encontró lo mismo en todos. Los estudios que acreditan su eficacia miden a quienes cumplen la dosis recomendada, media hora a la semana: con Khan Academy, 0,26 desviaciones típicas. Esa dosis la cumple alrededor del 5 % de los alumnos, que no son alumnos cualesquiera: lo que esos estudios miden es una asociación, no lo que pasaría si se obligara a todos a cumplirla. El artículo se titulaba con el hallazgo: estos programas benefician sobre todo a quien menos los necesita.²⁹

Estonia es el mejor país de Europa en PISA y ha hecho lo que recomiendan los manuales. En agosto de 2025 dio a sus 20.000 alumnos de 10.º y 11.º un asistente construido con investigadores propios, que ante la petición de una redacción responde preguntando para qué es y de cuánto

Figura 8 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 distintas, con poblaciones y denominadores distintos; ninguna mide directamente el esfuerzo. Lo que muestran son problemas de adopción y de continuidad, no un porcentaje común de alumnos que se esfuerzan.

tiempo se dispone. En marzo de 2026 habían activado la cuenta 7.700 alumnos, y de ellos la usaba cada semana el 47 %: menos de uno de cada cinco destinatarios. Antes del programa, entre el 64 y el 90 % de los alumnos estonios ya usaba herramientas comerciales para sus tareas.³⁰ La herramienta que obliga a pensar compite con la que da la respuesta, y el alumno ya tiene la segunda en el bolsillo.

Y el condado chino del capítulo 1 apunta en la misma dirección: de cada cinco alumnos con IA, uno mantiene el tiempo de trabajo.

4.3 Lo que dicen de sí mismos y lo que hacen

PISA pregunta a los alumnos por su actitud, y las respuestas españolas son buenas. El 68 % dice esforzarse más cuando la tarea se complica, frente al 60 % de la OCDE. El 43 % planifica cómo estudiar (37 %). El 62 % se hace preguntas para comprobar si ha entendido (49 %). Solo el 18 % cree que la escuela ha sido una pérdida de tiempo (24 %).¹⁵ Son los mismos alumnos que acaban de obtener el peor resultado de la serie. Entre lo declarado y lo hecho hay una distancia, y es el asunto de este capítulo.

La tendencia sí es informativa. En tres años, la curiosidad declarada baja 3,2 puntos en España y la perseverancia 3,3. En la OCDE, la proporción de alumnos a quienes aprender cosas nuevas no les aburre es del 57 %, y desde 2022 ha bajado en 60 sistemas educativos, en 30 de ellos más de diez puntos. No ha subido en ninguno. El apoyo que los alumnos dicen recibir de su familia cae en casi todos los países, sobre todo entre chicos y desfavorecidos, y hay más absentismo que en 2018. Donde más cae la lectura, más cae la implicación.^{10,13}

Sobre el tiempo de estudio conviene ser exacto. Los alumnos de la OCDE declaran 1,4 horas diarias de deberes. No es cierto, por tanto, que el alumno medio no dedique ni un minuto. Es cierto que ese tiempo es declarado, que ha bajado un 11 % en tres años y que una parte creciente consiste en copiar y pegar. El rendimiento sube con el tiempo de deberes hasta unas dos horas diarias, y quien dedica más tiempo muestra más curiosidad, más perseverancia y más metas propias en casi todos los países.¹³

4.4 Lo que compran las familias

El 47 % de los hogares españoles con hijos en edad escolar paga clases particulares. El gasto fue de 2.782 millones de euros en el curso 2023-2024, un 38 % más en términos reales que cuatro años antes, aunque casi todo ese aumento es rebote de la pandemia. El suelo está en el 30 % de los hogares más pobres y el techo por encima del 60 % en los más ricos.²⁷

El detalle interesante es otro. Lo que más crece son la primaria, los idiomas y las familias de renta baja. En ESO y Bachillerato, la participación ha bajado cinco puntos respecto a 2019-2020. Los autores no lo relacionan con la IA y los datos acaban en 2024. Hipótesis de este informe, que habrá que contrastar con la próxima encuesta: el chatbot se está comiendo la parte del negocio de la academia que consistía en explicar, que era la barata. La otra no la toca. Una explicación clara de cualquier tema de secundaria es hoy gratuita, instantánea e inagotable, y las familias siguen pagando casi tres mil millones al año. La encuesta no pregunta por qué pagan, y una parte compra idiomas o preparación de exámenes. La lectura de este informe es que, en el refuerzo escolar, lo que se paga

es sobre todo un adulto que espera al chico dos tardes por semana, mira lo que ha hecho y se da cuenta si no ha hecho nada.

De ahí sale la tesis central de este informe, que es una interpretación del conjunto y no el resultado de ningún estudio. En el aprendizaje de un adolescente corriente, el recurso escaso nunca fue la explicación. Es la obligación: alguien ante quien responder, una hora fijada, una mirada. La IA ha vuelto infinita la oferta de lo que ya sobraba y no produce ni un gramo de lo que falta.

No es pereza moral, y la economía del comportamiento lo explica sin sermones. Estudiar cuesta hoy y rinde dentro de dos años, y casi todo el mundo pesa de más el presente. La salida conocida es atarse las manos: un plazo, un testigo, una cita. En 2002, Dan Ariely y Klaus Wertenbroch lo midieron con universitarios: quienes tenían plazos impuestos desde fuera y repartidos por el curso rindieron más que quienes se los ponían a sí mismos, y estos más que quienes solo tenían la fecha final.³¹ Una academia, un preparador o un examen el jueves son eso: compromisos que otro hace cumplir. Un chatbot, por bueno que sea, está siempre disponible y nunca espera a nadie.

4.5 El tutor que no espera a que le pregunten

El mismo equipo de Tennessee ha probado la variante que se sigue de todo lo anterior. En otro programa de matemáticas, el asistente aparece sin que el alumno lo pida y no deja avanzar sin acertar. Los alumnos van más despacio, aciertan más y, una semana después y sin ayuda, rinden ligeramente mejor que el grupo de control.²⁸ Es poco y es preliminar. Señala la dirección: el diseño que funciona con el alumno que no quiere es el que no depende de que quiera. El fundador de Khan Academy lo formuló así tras el primer ensayo: la IA no podía estar al lado del contenido, tenía que ir dentro, de modo que el esfuerzo productivo fuera más difícil de esquivar.

4.6 Por qué estos veredictos

Los dos primeros ya ocurren, y el segundo no depende de la calidad del tutor: cuanto mejor sea el modelo, mejor hará también los deberes. Que un tutor voluntario mejore de forma apreciable al alumno medio de un país rico es «improbable» porque el ensayo más largo da entre 0,06 y 0,08 desviaciones típicas con un tutor que casi nadie usó, y nada en el diseño voluntario ataca el sesgo hacia el presente. El tutor integrado, que interviene sin que lo llamen, sube a «posible»: es el único diseño que quita la decisión de las manos del alumno, pero solo tiene indicios a una semana.

VEREDICTO Hasta 2031

Tutores de IA de buena calidad pedagógica, gratuitos o casi, al alcance de cualquier alumno	Muy probable Beneficioso en potencia; ya ocurre
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 los resultados del alumno medio de un país rico	Improbable Los ensayos largos dan entre nada y muy poco
Tutores integrados en el horario escolar, que intervienen sin que se les llame, con mejoras modestas y medidas	Posible Beneficioso; hay indicios, faltan ensayos largos

Los ensayos, uno por uno

EN BREVE

El metaanálisis que más se citaba para decir que ChatGPT mejora el aprendizaje ha sido retractado. Los ensayos serios son una docena y cuentan una historia coherente: la mejora aparece en la medida cercana —la práctica, el ejercicio de salida, el examen del día siguiente— y se encoge o desaparece en la lejana. Donde un tutor de IA ha funcionado con alumnos corrientes había un adulto al lado, un horario y una prueba sin la máquina; la excepción, Harvard, son alumnos a los que nadie tiene que obligar.

552.000

accesos al metaanálisis que atribuía a ChatGPT un efecto «grande» sobre el rendimiento ($g = 0,867$), retractado el 22 de abril de 2026³²

0,37

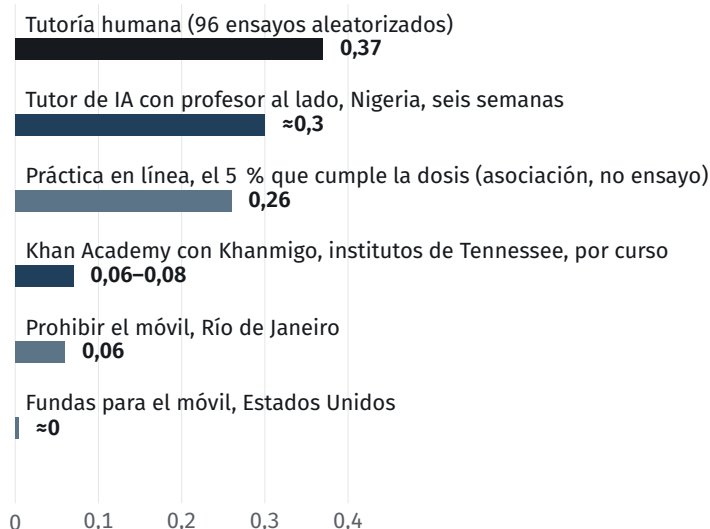
desviaciones típicas: efecto medio de la tutoría humana en 96 ensayos con sorteo; la vara con la que medir todo lo demás³³

0,06–0,08

efecto anual de la plataforma con tutor de IA en el ensayo más largo hecho en una escuela corriente²⁶

5.1 Un artículo retirado

En mayo de 2025, una revista del grupo Springer Nature publicó un metaanálisis de 51 estudios según el cual ChatGPT tenía un efecto grande sobre el rendimiento académico ($g = 0,867$) y moderado sobre el pensamiento de orden superior, y recomendaba extender su uso a todos los cursos. En dieciséis meses acumuló 552.000 accesos y 433 citas. Era la referencia de quien quería decir que la ciencia avalaba el chatbot en el aula.

Figura 9 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, duraciones, pruebas y diseños distintos: todas son estimaciones causales salvo la de práctica en línea, que es una asociación entre quienes cumplen la dosis. Compárese el orden de magnitud.

El 22 de abril de 2026, la revista lo retractó por «discrepancias en el metaanálisis» que minan la confianza en sus conclusiones. Los autores no respondieron. Lo habían denunciado dos investigadores de la Universidad Ártica de Noruega: errores al extraer los datos, mezcla de estudios no comparables, ningún filtro de calidad.³²

Siguen publicándose metaanálisis con efectos de esa talla, e incluso de más de una desviación típica. Un efecto así convertiría a un alumno mediano en uno del 16 % superior; ninguna intervención educativa seria ha logrado eso a gran escala. El problema es la materia prima: estudios de pocas semanas, con decenas de alumnos, que miden el resultado con la herramienta delante o con una prueba hecha por el propio investigador, y sin examen posterior sin IA.

5.2 La vara de medir

La promesa del tutor para cada niño tiene fecha. En 1984, el psicólogo Benjamin Bloom publicó que los alumnos con un tutor particular rendían dos desviaciones típicas por encima de los de una clase normal: el alumno medio con tutor superaba al 98 % de los demás. Lo llamó «el problema de las dos sigmas», porque nadie podía pagar un tutor para cada alumno, y es la cifra que citan quienes hoy venden tutores automáticos.³⁴ Los ensayos posteriores la han dejado en una quinta parte. El efecto medio de la tutoría humana, en 96 ensayos con sorteo, es de 0,37 desviaciones típicas, unos catorce puntos de percentil. Funciona mejor con profesores o monitores formados que con voluntarios, mejor dentro del horario escolar que después, y mejor en los primeros cursos.³³ Uno de los autores de ese metaanálisis es el mismo que ha dirigido el ensayo de Khanmigo. Sabe con qué compara.

5.3 Qué se ha ensayado

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

Estudio	Qué se hizo	Resultado	Pega principal
Turquía, instituto, ~1.000 alumnos (PNAS, 2025) ³	Práctica de matemáticas con un asistente libre, con uno que solo da pistas, o sin nada	Asistente libre: –17 % en el examen sin ayuda. Con pistas: ni daño ni mejora	Pocas sesiones
Harvard, Física, 194 alumnos (2025) ³⁵	Una lección con un tutor de IA en casa frente a la misma en clase activa	Más del doble de ganancia, en menos tiempo	Dos lecciones, prueba inmediata, alumnos de Harvard

continúa

Estudio	Qué se hizo	Resultado	Pega principal
Nigeria, Edo (Banco Mundial, 2025) ²⁵	Seis semanas de inglés por las tardes con IA y profesores presentes	≈0,3 d. t. en prueba en papel; también mejora el examen de fin de curso	Lo evalúa quien lo promueve; sin seguimiento largo
Reino Unido, 165 alumnos (Google y Eedi, 2025) ³⁶	Tutor de IA cuyos mensajes revisa un tutor humano, frente a tutor humano	66 % frente a 61 % de acierto en problemas nuevos	Pequeño; lo hace el fabricante; hay un humano en el circuito
EE. UU., 1.800 alumnos y 900 tutores (Stanford, 2024) ³⁷	La IA sugiere al tutor humano qué decir	+4 puntos de aprobados por lección; +9 con tutores flojos	Sin mejora detectable a fin de curso; solo dos meses de exposición
Tennessee, 18 institutos, dos cursos (NBER, 2026) ²⁶	Plataforma de práctica con tutor de IA voluntario, en horario lectivo	0,06–0,08 d. t. al año; el tutor apenas se usó	No separa el efecto del tutor del de la plataforma
OpenAI, «modo estudio», >300 universitarios (2026) ³⁸	40 minutos de estudio con el asistente guiado frente a buscador y vídeos	Microeconomía: ≈15 % más nota. Neurociencia: nada distinguible	Lo hace el fabricante; sin publicar; examen inmediato
EE. UU., adultos, cartas de presentación (2025) ⁹	Practicar redacción con IA o sin ella	Menos esfuerzo y mejor escritura posterior sin IA	Tarea breve, adultos motivados
China, 26.811 alumnos, 30 meses (CEPR, 2026) ¹	Uso libre, observado con fecha de adopción	–20 % en exámenes; –24 y –18 % en los de acceso	Un condado; adopción declarada

continúa

Estudio	Qué se hizo	Resultado	Pega principal
EE. UU., plataforma ALEKS, 10 años (2026) ²	Uso libre, deducido de la huella en los tiempos	–31 % de tiempo; peor retención bajo vigilancia	No observa la IA; significación justa

5.4 La regularidad: la mejora vive en la medida cercana

Puestos en fila, los estudios dibujan algo que ninguno dice por separado. En Turquía, el asistente con pistas dispara los ejercicios resueltos durante la práctica y no mueve el examen. En el ensayo de Stanford, los alumnos aprueban más ejercicios de salida de cada lección y no se detecta mejora en la prueba estandarizada de fin de curso, con solo dos meses de exposición. En Harvard, la ganancia se mide al acabar la lección. En el de OpenAI, tras cuarenta minutos de estudio. En China y en la plataforma estadounidense, donde se mide lejos y el uso era libre, el signo es negativo. Hay dos excepciones, y no parecen casuales: en Nigeria mejora también el examen ordinario de fin de curso, y en Tennessee queda un efecto pequeño al cabo de dos cursos. En las dos había horario, adulto y examen sin máquina.

Hipótesis de este informe: con IA, cuanto más cerca del uso se mide, mejor sale, y la mayor parte de la literatura mide muy cerca. No es fraude, es lo barato: un examen a la semana siguiente sin la herramienta cuesta poco, y uno al año siguiente cuesta un proyecto. Pero el orden de los resultados es el que predice la psicología del aprendizaje: Bjork distingue el rendimiento durante la práctica de lo que queda después, que son cosas distintas y a veces opuestas.⁸ Es una hipótesis y tiene un límite claro: los estudios comparan intervenciones distintas —uso libre, programa vigilado, IA detrás del tutor—, de modo que entre ellos no cambia solo la distancia de la medida. Para comprobarla haría falta seguir un mismo tratamiento y examinar a distintas distancias.

OpenAI lo admite en el texto donde presenta su propio ensayo: lo que importa es si las ganancias «siguen ahí con el tiempo», y los métodos habituales no ven si una mejora del recuerdo a corto plazo viene acompañada

de pérdidas en perseverancia, motivación propia o resolución creativa de problemas. Ha anunciado una batería de medida a largo plazo con las universidades de Tartu y Stanford.³⁸ Que el vendedor reconozca que no sabe si su producto enseña a medio plazo es el dato más honrado del año.

5.5 Donde sí ha funcionado

Nigeria es el mejor resultado, y conviene mirar por qué. Seis semanas, por las tardes, en la escuela, con profesores presentes que guiaban la sesión; una prueba final en papel, sin IA; y efecto también en los exámenes ordinarios de fin de curso, que trataban de otra materia. Cada día más de asistencia mejoraba el resultado, y las chicas, que partían por detrás, ganaron más. Los autores lo equiparan a casi dos cursos de aprendizaje ordinario y lo sitúan por encima del 80 % de las intervenciones evaluadas con sorteo en países en desarrollo.²⁵ Tenía todo lo que falta en el dormitorio de un adolescente europeo: un sitio, una hora, un adulto y un examen sin máquina. Y partía de un listón distinto: allí la alternativa no era un profesor particular, sino ninguna.

El segundo caso sólido es el menos vistoso. En el ensayo de Stanford, la IA no habla con el alumno: le sopla al tutor humano, en directo, qué preguntar. Los alumnos de los tutores peor valorados pasan del 56 al 65 % de lecciones superadas y alcanzan a los de los tutores buenos sin ayuda. La herramienta cuesta veinte dólares por tutor y año; los tutores van aparte. Incluso ahí, solo se usó en el 29 % de las sesiones.³⁷

5.6 Cinco preguntas para leer el próximo titular

1. ¿Se midió sin la herramienta? Si el examen se hizo con la IA delante, mide a la IA.
2. ¿Cuánto tiempo después? Un examen sin ayuda el mismo día dice que algo se ha adquirido, no que vaya a durar. Retener se comprueba a la semana o al mes, y saber usarlo en un problema distinto, más tarde todavía. En el estudio chino, la pérdida en los exámenes de acceso siguió creciendo dos años.

3. ¿Quién decidía usarla? Si era voluntario, hay que mirar cuántos la usaron y cómo antes de mirar el efecto.
4. ¿Comparado con qué? Con nada, casi todo funciona. Con un profesor o un tutor humano, casi nada.
5. ¿Quién lo hace y quién lo paga? Tres de los diez estudios de la tabla los ha hecho el fabricante o alguien de la casa.

5.7 Por qué estos veredictos

Habrán más metaanálisis inflados porque la materia prima no ha cambiado: estudios cortos, pequeños y medidos con la herramienta delante. Que un tutor voluntario iguale a la tutoría humana en ensayos largos es «muy improbable»: la tutoría humana funciona, según este informe, por la obligación que impone, y eso es justo lo que el tutor voluntario no tiene. La IA detrás del adulto es «probable» porque ya tiene dos ensayos con sorteo a favor y cuesta veinte dólares. Los efectos grandes donde faltan profesores se quedan en «posible»: hay un solo ensayo bueno, evaluado por quien lo promueve.

VEREDICTO Hasta 2031

Nuevos metaanálisis con efectos enormes de la IA sobre el aprendizaje, y nuevas retractaciones

■ ■ ■ ■ ■ **Muy probable**
Limitado; contamina el debate

Ensayos largos, con examen sin IA, que muestren efectos de un tutor de IA voluntario iguales a los de la tutoría humana

■ ■ ■ ■ ■ **Muy improbable**
—

Efectos modestos y repetidos cuando la IA ayuda al profesor o al tutor humano

■ ■ ■ ■ ■ **Probable**
Beneficioso; barato

Efectos grandes de tutores de IA con adulto presente en países con escasez de profesores

■ ■ ■ ■ ■ **Posible**
Muy beneficioso; un solo ensayo bueno hasta hoy

Milagros y marchas atrás

EN BREVE

El milagro salvadoreño son siete alumnos voluntarios por escuela, en escuelas elegidas por estar bien equipadas y dispuestas, sin medida anterior ni grupo de comparación. La misma semana, el examen de verdad daba 346 puntos en matemáticas. Corea se gastó 850 millones de dólares en libros de texto con IA y les quitó el rango de libro de texto a los cuatro meses. China, que es quien más empuja la IA, prohíbe el chatbot abierto a los niños de primaria. Y la bala de plata anterior, prohibir el móvil, ha dado entre nada y muy poco. Casi nada de lo que se anuncia en rueda de prensa se ha medido antes.

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 centros³⁹

346

puntuación de El Salvador en matemáticas en PISA 2025, con 7.073 alumnos de 278 centros: tres puntos más que en 2022 y 117 menos que la media de la OCDE⁴⁰

≈0

efecto medio sobre las notas de guardar el móvil en una funda con cierre, en el mayor estudio hecho en Estados Unidos⁴¹

6.1 El Salvador

En diciembre de 2025, el Gobierno salvadoreño y la empresa xAI anunciaron que llevarían su asistente como tutor a más de 5.000 escuelas públicas y a más de un millón de alumnos en dos años. En septiembre de 2026, el presidente y el Banco Mundial presentaron los resultados de una prueba

hecha en junio en 171 escuelas del programa piloto de modernización: niveles comparables a los de Alemania y Suecia.⁴²

Conviene leer lo que se midió. La prueba fue PISA para Centros Educativos, una herramienta que la OCDE ofrece a escuelas sueltas; no es el estudio PISA. La hicieron 1.198 alumnos de 171 escuelas, siete por escuela de media, y eran voluntarios. Las escuelas no se eligieron por sorteo: según el propio Gobierno, entraron en el piloto por sus condiciones de infraestructura, su acceso a internet y la disposición de su comunidad educativa. No hay medida anterior ni grupo de comparación. Y el piloto no consistía en un chatbot: incluía dispositivos, conexión, plataformas, un método de enseñanza estructurada, materiales y acompañamiento a profesores y directores.³⁹

Los responsables fueron más prudentes que los titulares. La ministra de Educación dijo que no se trataba de una muestra nacional ni de una medición del sistema público. El presidente, que los datos no permiten concluir que el sistema salvadoreño esté al nivel de Alemania. El Banco Mundial, que los resultados valen solo para las escuelas participantes.

Una diferencia de 130 puntos equivale a seis cursos. Ninguna intervención conocida produce eso en unos meses. La sospecha razonable —que se midió a los mejores de las mejores— se queda corta: se midió a siete voluntarios por escuela, en escuelas elegidas por estar preparadas.

Lo notable es que El Salvador sí tiene un dato bueno, y nadie lo ha usado. En el PISA de verdad, con 7.073 alumnos de 278 centros, obtuvo 366 puntos en lectura, 346 en matemáticas y 385 en ciencias: dos, tres y doce más que en 2022, en un ciclo en que casi todo el mundo baja.⁴⁰ Y lo hizo mientras escolarizaba a más jóvenes: es uno de los tres únicos países que en una década han ampliado mucho la parte de la población que se examina y han mejorado a la vez.¹² Sigue a más de cien puntos de la media de la OCDE, y la mejora es anterior al acuerdo con xAI. Es un progreso real, lento y sin relación demostrada con ningún chatbot.

6.2 Corea del Sur

Corea puso en marcha en marzo de 2025 libros de texto digitales con IA para matemáticas, inglés e informática, con una inversión de unos 850

millones de dólares. La adopción pasó del 37 al 19 % de los centros en un semestre. En agosto, el Parlamento les retiró la condición de libro de texto: son material complementario, y el centro que los quiera se los paga. A mediados de octubre, más de la mitad de las 4.095 escuelas adheridas se había dado de baja. Las quejas de familias y profesores fueron errores de contenido, datos de menores, más horas de pantalla y más trabajo para el docente.⁴³ Corea pierde quince puntos en lectura en PISA 2025.

6.3 Estonia

Estonia ha hecho lo contrario y da menos titulares⁴⁵. Un programa de cuatro millones de euros a tres años, a medias entre el Ministerio y varias empresas, empezó en agosto de 2025 por los dos primeros cursos de secundaria superior, con unos 4.900 profesores implicados y una herramienta propia que pregunta en lugar de responder. Ha encargado a las universidades de Tartu y Stanford un estudio de impacto que todavía no tiene resultados, y no ha anunciado ninguno.³⁰ Lo único que se sabe es el uso, que es bajo, como se vio en el capítulo 4.

6.4 China

El país que ha declarado la IA prioridad nacional publicó en mayo de 2025 dos guías para la escuela. Una hace obligatoria la enseñanza sobre IA desde primaria; Pekín exige al menos ocho horas al año desde los seis años. La otra gradúa el uso por edades: en primaria, los alumnos no pueden usar por su cuenta herramientas generativas abiertas; en los primeros cursos de secundaria pueden explorar cómo razonan; en bachillerato, el uso es más amplio. Está prohibido entregar como deberes contenido generado, cada centro debe tener una lista de herramientas autorizadas, y los profesores no deben usarlas para calificar.⁴⁴

Es la regulación escolar más restrictiva de un país grande, y viene de quien menos sospecha despierta de tecnofobia. Llegó tarde para el condado del capítulo 1: cuando se publicó, cuatro de cada cinco alumnos ya usaban estas herramientas en casa. La guía se dirige también a las

familias, a las que pide supervisar y coordinarse con la escuela.⁴⁵ Cuánto cambia eso lo que pasa en un dormitorio, todavía no lo ha medido nadie.

6.5 La bala de plata anterior

Antes de la IA, la solución era prohibir el móvil, y ya hay medidas serias. El mayor estudio, con datos de todo Estados Unidos sobre las fundas con cierre que usan miles de centros, encuentra que reducen mucho el uso del teléfono, que el primer año aumentan los incidentes de disciplina y empeora el bienestar, que después ambos efectos se corrigen, y que el efecto sobre las notas es «sistemáticamente cercano a cero», con algo positivo en los institutos y algo negativo en los primeros cursos de secundaria.⁴¹ En Florida, las notas mejoran a partir del segundo año y bajan las faltas sin justificar.⁴⁶ En Río de Janeiro, la prohibición de 2023 sube las notas 0,06 desviaciones típicas.⁴⁷ PISA añade que los países que más han prohibido desde 2022 no han evolucionado mejor.¹³

Prohibir el móvil reduce la distracción y mueve las notas entre nada y un poco, según la edad, el cumplimiento y los años que lleve la medida. Puede merecer la pena, y no era la palanca. Con el chatbot hay una razón añadida para no esperar mucho de prohibirlo en el centro: el daño medido ocurre en casa.

6.6 Por qué estos veredictos

Los anuncios sin grupo de comparación seguirán porque salen gratis: el titular llega el primer día y la corrección, si llega, meses después. Otra marcha atrás cara es «probable» porque los despliegues nacionales se están decidiendo antes de medir, que es exactamente lo que hizo Corea. Las restricciones por edad son «probables» en Europa porque China ha dado el ejemplo y porque son baratas de aprobar; su límite es que el daño medido ocurre en casa. Un programa nacional con mejora demostrada en un examen externo es «improbable»: 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 a familias y
gobiernos

Algún país da marcha atrás en un
despliegue masivo de IA escolar tras
gastar cientos de millones

■■■■■ **Probable**
Grave para sus presupuestos; ya
ha ocurrido una vez

Restricciones por edad al uso de chatbots
abiertos en la escuela en varios países
europeos

■■■■■ **Probable**
Limitado; afecta al centro, no a la
casa

Un programa nacional de IA educativa
con mejora demostrada en un examen
externo

■■■■■ **Improbable**
Beneficioso; Estonia es la
candidata

El profesor

EN BREVE

Más de la mitad de los alumnos españoles usa la IA cada semana para estudiar; en 2024 la usaba en su trabajo un tercio de sus profesores, y para otra cosa. El uso con mejor evidencia de toda esta literatura no pone la IA delante del alumno, sino detrás del adulto: le prepara la clase y le sopla al tutor qué preguntar. Lo que la IA obliga a cambiar al profesor no es la explicación, que sigue valiendo, sino la evaluación: qué cuenta para la nota y dónde se hace.

**35 % y
54 %**

profesores españoles de secundaria que usaban IA en su trabajo en 2024, y alumnos españoles de 15 años que la usan cada semana para estudiar en 2025; encuestas y preguntas distintas^{15, 48}

13 → 16 %

parte de la clase que se va en mantener el orden, 2018-2024, media de la OCDE¹²

+9 puntos

lecciones superadas por los alumnos de los tutores peor valorados cuando una IA les sugiere qué decir; coste de la herramienta, 20 dólares por tutor y año³⁷

7.1 Quién usa qué

En la encuesta de la OCDE al profesorado, hecha en 2024, usa IA en su trabajo el 35 % de los docentes españoles de secundaria y el 28 % de los de primaria. La media de la OCDE es el 36 %; en Singapur y los Emiratos, el 75 %; en Francia, el 14 %; en Japón, el 17 %. La usan para ponerse

al día o resumir un tema (68 %) y para preparar clases (64 %); uno de cada cuatro, para corregir. De los que no la usan en España, tres de cada cuatro dicen que no tienen formación para hacerlo. Apenas cuatro de cada diez le ven beneficios, por debajo de la media, y en la OCDE el 72 % cree que daña la integridad académica, porque permite presentar como propio lo ajeno.^{4, 48}

Un año después, el 54 % de los alumnos decía usarla cada semana para estudiar. Las dos encuestas no preguntan lo mismo ni el mismo año, de modo que la comparación es solo orientativa. Lo que sí consta es que quien tiene que rediseñar los deberes y los exámenes dice no estar formado para ello, y nadie le ha dado tiempo: el profesor español declara más estrés por tareas administrativas que la media (64 % frente a 52 %), valora peor la formación que recibe (la considera de calidad el 56 %, frente al 77 %) y solo el 39 % dice que esa formación le ha servido para gestionar el comportamiento en el aula, frente al 58 %.⁴⁸ Y el 49 % de los alumnos españoles está en centros cuyo director dice que la falta de profesorado dificulta la enseñanza; en 2022 era el 41 %, y la media de la OCDE es el 39 %.¹⁵

7.2 Detrás del adulto, no delante del alumno

El ensayo de Stanford citado en el capítulo anterior es el uso mejor documentado de toda esta literatura. Novecientos tutores y 1.800 alumnos de primaria y secundaria, casi todos de familias con pocos recursos, en sesiones de refuerzo de matemáticas dentro del horario escolar. A la mitad de los tutores, elegidos por sorteo, un asistente les proponía en directo qué decir: una pregunta guía, una pista, otra manera de explicar. El alumno no veía la IA. Los tutores con asistente preguntaron más y dieron menos veces la respuesta. La probabilidad de que sus alumnos superaran la lección subió cuatro puntos; con los tutores peor valorados, nueve, del 56 al 65 %, hasta igualar a los alumnos de los buenos tutores.³⁷

Tiene límites: dos meses, un distrito, y ningún efecto detectable en el examen de fin de curso. Pero el diseño resuelve los dos problemas de los capítulos anteriores. La obligación la pone el adulto, y la IA aporta lo que al adulto inexperto le falta, que es oficio. La síntesis de la OCDE la

destaca como la evidencia más robusta que hay.⁴ Lo mismo vale para el ensayo británico en el que un tutor humano revisaba cada mensaje de la IA antes de que llegara al alumno: aprobó tres de cada cuatro sin tocarlos, y los alumnos rindieron al menos igual que con el tutor humano solo.³⁶

7.3 Lo que cambia es la evaluación

Los autores del estudio chino terminan con recomendaciones para profesores y padres, y ninguna es técnica: vigilar el tiempo que de verdad se dedica a los deberes, que dice más que su nota, y comprobar lo trabajado en casa con pruebas cortas y frecuentes en clase.¹ Los del estudio israelí piden mantener exámenes presenciales junto a trabajos rediseñados, y avisan a los empleadores de que la nota informa menos.⁵ La OCDE propone un orden para introducir la herramienta: primero aprender sin IA generativa, después con IA educativa, y solo después con IA general.⁴

Ese orden es el de la calculadora. El metaanálisis clásico sobre ella, con 54 estudios, concluyó que no había perjudicado las destrezas matemáticas, y que ayudaba cuando formaba parte de la enseñanza y del examen.⁴⁹ Quien recuerda que con la calculadora también se anunció el desastre tiene razón en eso. Se equivoca si cree que es la misma herramienta. La calculadora se queda con un paso estrecho, la aritmética, y deja al alumno el planteamiento. El chatbot se queda con el planteamiento, el desarrollo y la redacción. Y la calculadora entró en un currículo que decidió cuándo: primero a mano, luego con máquina.

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 a cualquier hora. Lo que ninguno hace es conseguir que treinta adolescentes estén cincuenta minutos intentando algo difícil, ni saber cuál de ellos no ha entendido. Eso es orden en el aula, relación y examen, las tres cosas que TALIS y PISA dicen que más se han deteriorado.

7.4 Por qué estos veredictos

Que la mayoría del profesorado use IA para preparar clases es «muy probable»: ya lo hace un tercio, ahorra tiempo y no tiene coste para el aprendizaje. El desplazamiento de la nota hacia pruebas presenciales es «probable» porque es la única respuesta que no depende de detectar el texto generado; lo frena que cuesta horas de corrección que nadie ha presupuestado. El copiloto de tutores es solo «posible» porque exige antes lo caro, que son los tutores. Sustituir profesores por tutores de IA es «muy improbable» por la tesis de este informe: el profesor nunca fue, sobre todo, una fuente de explicaciones.

VEREDICTO Hasta 2031	
La mayoría del profesorado usa IA para preparar clases y materiales	Muy probable Beneficioso; ahorra tiempo
Los trabajos hechos en casa pierden peso en la nota a favor de pruebas presenciales, orales y en papel	Probable Beneficioso; exige horas de corrección
IA como copiloto de tutores y profesores noveles, con efectos modestos y medidos	Posible Beneficioso y barato
Sustitución de profesores por tutores de IA en la enseñanza obligatoria de un país rico	Muy improbable Grave; nada lo respalda

España

EN BREVE

España va a gastar cerca de 2.800 millones de euros al año en bajar el número de alumnos por aula, una medida sin efecto demostrado sobre el aprendizaje. Es la misma cifra que sus familias gastan en clases particulares, el pariente privado de lo que mejor evidencia tiene, que es la tutoría. Tiene la mitad de alumnos excelentes que la media, ha prohibido ya el móvil, sus alumnos usan la IA más que la media y sus profesores dicen no estar formados. El sitio donde la IA tendría sentido económico es el que nadie ha propuesto: detrás de un tutor, en horario escolar.

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 ESO⁵⁰

2.782 M€

lo que gastan al año las familias españolas en clases particulares²⁷

**4 %, 4 % y
2 %**

alumnos españoles excelentes en ciencias, matemáticas y lectura; en la OCDE, 7, 8 y 6 %¹⁵

8.1 El retrato

Reunido lo que dicen los capítulos anteriores, España es un caso particular. No tiene más alumnos por debajo del mínimo que la media de la OCDE, y sí la mitad por arriba. Su desigualdad por origen social es menor que la media. Sus alumnos declaran más perseverancia y más sentido de pertenencia al centro que casi nadie, y obtienen el peor resultado de su serie. Ha prohibido el móvil en el recinto escolar para el 86 % del

alumnado, casi el doble que la media, sin que se note en las notas. Sus alumnos usan la IA para estudiar más que la media en todas las finalidades que mide PISA. Sus profesores la usan como la media, la valoran peor y dicen no estar formados. Casi la mitad de los alumnos está en centros a los que, según su director, les faltan profesores, y siete de cada diez en centros a los que les falta personal de apoyo, frente al 39 % de la OCDE.^{15,48}

Y uno de cada cuatro alumnos de 15 años no está en el curso que le corresponde.¹¹

8.2 La apuesta: bajar la ratio

El 11 de noviembre de 2025, el Ministerio presentó el anteproyecto de ley que baja el máximo de alumnos por aula de 25 a 22 en primaria y de 30 a 25 en ESO, y cuenta doble a los alumnos con necesidades especiales. El centro de políticas económicas de Esade ha calculado lo que cuesta. El coste presupuestario neto será, de media, de al menos 2.818 millones de euros al año entre 2027 y 2036. En 2031, con la ley del todo aplicada, 5.169 millones: el 13,8 % de todo el gasto público en primaria y ESO, y eso después de restar lo que ahorra la caída de la natalidad. Andalucía, Cataluña y Madrid soportan el 63 %. Y es regresiva: como las aulas llenas están en zonas urbanas con mucha demanda, el 31 % del gasto va a los centros con alumnado de nivel social más alto y el 22 % a los de nivel más bajo.⁵⁰

Sobre el efecto, el mismo informe recoge el metaanálisis más completo que existe, de 2025, con 66 estudios causales: reducir el tamaño de la clase no tiene efecto agregado significativo sobre el aprendizaje, la repetición ni el bienestar del alumno. Que no se detecte un efecto medio no prueba que sea cero en todas partes, y por eso los autores no proponen abandonar la medida, sino acotarla. Mejora algo la satisfacción del profesor, menos de lo que la mejoraría el mismo dinero puesto en su sueldo. Los autores recomiendan limitar la medida a los centros difíciles y llevar el resto del dinero a lo que sí tiene evidencia: tutoría intensiva en grupos pequeños, complementos para quien enseña en centros complicados y un sistema de iniciación a la profesión.

8.3 Lo que hacen las familias

Las familias ya han votado con la cartera, como se vio en el capítulo 4: casi la mitad paga clases particulares, por 2.782 millones al año. La cifra es casi idéntica a la de la ley. El Estado se dispone 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.

8.4 Donde la IA tendría sentido

Los programas públicos de refuerzo en grupo pequeño existen en España y tienen el problema de siempre: cuestan, y faltan tutores con oficio. El ensayo de Stanford aborda justo ese cuello de botella. Un asistente que sugiere al tutor qué preguntar cuesta veinte dólares por tutor y año, y sus efectos son mayores cuanto menos experiencia tiene el tutor.³⁷ El asistente es la parte barata. Lo caro son los tutores, y ese es exactamente el dinero que la ley de ratios destina a otra cosa. Un programa nacional de refuerzo en horario escolar, con tutores apoyados por IA y evaluado con sorteo antes de extenderlo, es el tipo de gasto al que el propio informe de Esade propone llevar parte de ese dinero. Con la IA dentro, nadie lo ha propuesto. El debate español sobre IA y escuela se reduce a si se permite o se prohíbe el chatbot en clase, que es donde menos daño hace.

8.5 Las comunidades

Educación es competencia autonómica, y el mapa es desigual. Madrid, Castilla y León y Asturias rinden por encima de la media de la OCDE en ciencias; Ceuta, Melilla y la Comunidad Valenciana quedan muy por debajo; Cataluña no tiene dato publicable por haber excluido a casi uno de cada cuatro alumnos de la muestra.^{15, 16} Varias comunidades han restringido las pantallas en primaria y algunas han empezado pilotos con asistentes; no consta ninguna evaluación publicada con grupo de comparación.

8.6 Los que llegan ahora

Los alumnos examinados en PISA 2025 nacieron en 2009. Tenían trece años cuando apareció ChatGPT y han hecho con él la segunda mitad de la ESO. Los que se examinen en el próximo PISA, el de 2029, nacieron en 2013: tenían nueve años, estaban en 4.º de primaria y habrán hecho con él toda la ESO. Si el mecanismo del capítulo 1 actúa en España como en aquel condado chino, lo razonable es esperar que la nota de los deberes y de los trabajos siga subiendo, que los suspensos bajen y que el examen externo empeore. España tiene pocos exámenes externos: la prueba de acceso a la universidad, las evaluaciones de diagnóstico y PISA. Son los únicos termómetros que la IA no ha roto.

8.7 Por qué estos veredictos

El primero es «probable» porque el anteproyecto está presentado y el mejor metaanálisis no encuentra efecto medio. El segundo, porque es lo ya medido en China y en Israel, y España tiene pocos exámenes externos que lo delaten. El programa de tutoría con IA de apoyo no está en la agenda de nadie. Y recuperar en 2029 el nivel de 2015 serían 45 puntos de lectura en cuatro años: ningún país ha recuperado eso en un ciclo.

VEREDICTO Hasta 2031

Ley de ratios aprobada, sin mejora apreciable en evaluaciones externas

■ ■ ■ ■ ■ Probable

Grave: unos 2.800 M€ al año

Suben las notas internas y bajan las de las pruebas externas

■ ■ ■ ■ ■ Probable

Grave; oculta el problema

Programa nacional de tutoría escolar con IA de apoyo, evaluado con sorteo

■ ■ ■ ■ ■ Improbable

Beneficioso; hoy no está en la agenda

España recupera en PISA 2029 el nivel de 2015

■ ■ ■ ■ ■ Muy improbable

—

Cómo se hace aprender a quien no quiere

EN BREVE

Ningún chatbot ha demostrado motivar a quien no quiere. Lo que hace trabajar a un adolescente sin ganas es lo de siempre: una hora fijada, un adulto que espera el resultado y un examen que no se puede delegar. La IA sirve dentro de ese marco y estorba fuera de él. El diseño que funciona con el alumno que no quiere es el que no depende de que quiera.

9.1 Lo que dicen juntos los capítulos anteriores

1. Con acceso libre, cuatro de cada cinco alumnos usan la IA para acabar antes, y pierden. El quinto mantiene el tiempo de trabajo y no pierde nada.
2. El tutor que no da la respuesta existe y es gratuito. En los programas con datos, el uso real es bajo o superficial: de una vigésima parte a un tercio, según la medida.
3. Donde un tutor de IA ha funcionado con alumnos corrientes había un sitio, una hora, un adulto presente y una prueba sin máquina. La excepción, Harvard, son alumnos a los que nadie tiene que obligar.
4. El uso con mejor evidencia es la IA detrás del adulto, no delante del alumno.
5. La mejora que se mide el mismo día no garantiza que dure. La que se mide al cabo de un curso suele ser pequeña.

6. Prohibir el dispositivo en el centro reduce la distracción y mueve las notas poco o nada. El daño medido de la IA ocurre en casa.

De ahí salen seis principios. Ninguno es nuevo, y eso es una ventaja: todos tenían evidencia antes de que existiera ChatGPT.

9.2 Uno: la obligación la pone una persona

La tutoría humana rinde 0,37 desviaciones típicas de media, y más cuando se hace dentro del horario escolar que después.³³ El metaanálisis no aísla por qué funciona. La lectura de este informe es que 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. Todo plan para el alumno sin ganas empieza por decidir quién es esa persona y a qué hora. Si la respuesta es «el propio alumno, cuando le parezca», el plan es para el otro 20 %.

9.3 Dos: el examen sin máquina, corto y frecuente

Intentar recordar algo sin mirarlo enseña más que releerlo. Es uno de los resultados más sólidos de la psicología del aprendizaje: en el experimento clásico de Henry Roediger y Jeffrey Karpicke, quienes se examinaron recordaban más al cabo de una semana que quienes habían vuelto a estudiar el texto el mismo número de veces.⁵¹ Casi nadie estudia así por su cuenta, porque resulta incómodo y da la sensación de saber menos.⁷ Una prueba de diez minutos en papel, cada semana, sobre los deberes de esa semana hace tres cosas a la vez: obliga a recuperar, devuelve al profesor el termómetro que la IA ha roto y convierte los deberes copiados en un mal negocio para el alumno, que es el único argumento que entiende quien no quiere. No hace falta que cuente mucho para la nota. Hace falta que ocurra siempre.

9.4 Tres: mirar el tiempo, no la nota

En el estudio chino, la nota de los deberes de quien usa IA no dice nada, o dice lo contrario. El tiempo sí: quien mantiene el tiempo aprende lo mismo que antes.¹ Para un padre es la señal más barata que hay. Si una hoja de problemas que llevaba una hora se entrega perfecta en quince minutos, ha pasado algo. Para un centro, las plataformas de deberes ya registran la duración. Los autores avisan de que obligar a dedicar más tiempo no garantiza recuperar lo perdido, porque quien elige trabajar también es distinto. Pero como señal de alarma, el tiempo funciona y la nota ya no.

9.5 Cuatro: primero sin, luego con

El orden que propone la OCDE —sin IA, después con IA educativa, después con IA general— es el de la calculadora y el de cualquier oficio: se aprende a hacerlo a mano antes de delegarlo, porque solo quien sabe hacerlo distingue un resultado bueno de uno malo.⁴ China lo ha convertido en norma por edades. La consecuencia práctica es que la pregunta «¿se permite la IA?» está mal planteada. La pregunta es en qué tareas, a qué edad y después de haber demostrado qué.

9.6 Cinco: si hay tutor de IA, que no espere a que le llamen

El tutor voluntario fracasa porque depende de que el alumno pida ayuda, y el alumno sin ganas no la pide. Las dos variantes con indicios a favor quitan esa decisión de sus manos. Una es el asistente integrado en el ejercicio, que interviene cuando detecta el error y no deja avanzar sin acertar: los alumnos van más despacio, que es justo lo que se busca, y rinden algo mejor una semana después.²⁸ La otra es la sesión fijada con un adulto en la sala, como en Nigeria.²⁵ Con alumnos corrientes, que el tiempo por problema se desplome es mala señal: es lo que midió la

plataforma estadounidense del capítulo 1. No es una ley —en Harvard aprendieron más en menos tiempo—, pero allí nadie pegaba el enunciado.

Hay un uso individual que sí tiene respaldo: ver un ejemplo bien resuelto y, después, resolver otro sin ayuda.⁹ La segunda mitad no es opcional.

9.7 Seis: la IA, detrás del adulto

Preparar materiales, generar variantes de un ejercicio, proponer al tutor la siguiente pregunta, corregir un primer borrador para que el profesor revise. Es donde están los efectos medidos, el coste es mínimo y no compete con el esfuerzo del alumno.^{36, 37}

Tabla 2 Qué funciona con el alumno que no quiere, y qué no

Medida	Qué se sabe	Para quién sirve
Dar acceso a un tutor de IA voluntario	Se usa poco; efecto pequeño	Para el alumno que ya quería
Prohibir el móvil en el centro	Menos distracción; notas entre iguales y algo mejores. Del chatbot no hay ensayos	Para el ambiente del aula
Deberes en casa que cuentan para la nota	La nota sube y deja de informar	Para nadie
Prueba corta semanal en papel	Evidencia sólida anterior a la IA	Para todos; devuelve la medida
Tutoría humana, individual o en grupo pequeño	0,37 d. t. de media en 96 ensayos; más efecto en horario escolar	Sobre todo para el que va mal
IA que apoya al tutor o al profesor	Efectos modestos y repetidos; barata	Para el tutor con poco oficio
Tutor de IA integrado, que interviene solo	Indicios a una semana; faltan ensayos largos	Para el alumno medio, en clase

9.8 Lo que no va a pasar

A juicio de este informe, no va a aparecer un asistente tan bueno que el alumno sin ganas lo prefiera al que le da la respuesta. Los autores del estudio chino lo dicen con todas las letras: el problema de incentivos no desaparecerá con los avances de la técnica. Cuanto mejor sea el modelo, mejor hará los deberes. La respuesta es de organización: qué se hace en clase, qué cuenta para la nota, quién mira.

Y no va a bastar con enseñar «uso responsable». PISA encuentra que quienes han aprendido en clase a evaluar lo que genera una IA rinden algo mejor, y merece la pena hacerlo. Pero un adolescente que sabe perfectamente que copiar no enseña copia igual si nada depende de ello. Saber lo que conviene nunca ha bastado a los quince años. Tampoco a los cuarenta.

9.9 Por qué estos veredictos

El primero es la hipótesis central de este informe y está por ensayar: el examen frecuente sin máquina ataca a la vez la medida rota y el incentivo a copiar, y el efecto de examinarse es de los mejor replicados de la psicología. El segundo es «muy improbable» porque el problema no es de calidad, sino de incentivos. El tercero es «probable» porque hacia ahí se mueven ya las plataformas tras el ensayo de Tennessee.

VEREDICTO Hasta 2031

Los centros que pasen la evaluación a pruebas presenciales frecuentes conservan el nivel mejor que los que no

■ ■ ■ ■ ■ Probable

Beneficioso; es la hipótesis central de este informe y está por ensayar

Un asistente tan atractivo que el alumno medio lo prefiera, por sí solo, al que da la respuesta

■ ■ ■ ■ ■ Muy improbable

—

Tutores de IA integrados en plataformas escolares, de uso obligatorio en clase, generalizados en países ricos

■ ■ ■ ■ ■ Probable

Beneficioso si se mide sin la herramienta

Qué hacer

EN BREVE

La escuela no se hunde y el chatbot no la salva. Lo que funciona tiene poco de técnico: hacer primero a mano, comprobar sin ayuda una semana después, fijarse en el tiempo y no en la nota, y que haya un adulto que espere el trabajo. Casi todo se puede hacer desde mañana, sin esperar a ninguna ley.

10.1 Tres conclusiones con las que trabajar unos años

La IA no originó la caída de las notas: la encontró empezada. Lo más probable es que la acelere por un sitio concreto, los deberes. Tratarla como la causa de todo lleva a prohibirla en clase, que es donde menos daño hace, y a no tocar lo que sí falla.

El problema nunca fue la falta de buenas explicaciones, y ahora sobran. Lo escaso es un motivo para sentarse. Eso lo da una persona, un horario y un examen, y ningún programa.

La nota de lo hecho en casa ha dejado de informar. Quien tenga que decidir algo sobre un alumno —él mismo, sus padres, su profesor, una universidad— necesita una medida tomada sin máquina.

10.2 Si estudia: bachillerato, universidad, una oposición

La prueba de la semana siguiente. Una semana después de trabajar un tema con ayuda de una IA, intente un ejercicio parecido sin nada delante. Si no sale, aquello no lo sabe, por bien que quedara la hoja. Es el único control que no engaña, y cuesta diez minutos.

Separe las sesiones de producir de las de aprender. Para entregar algo que no necesita dominar, la herramienta es legítima y rápida. Para lo que entra en el examen, el orden es el contrario: primero el intento propio, con errores, y después la consulta.

Pregunte como a un profesor, no como a un solucionario. Pida que le examine, que encuentre el fallo en su razonamiento, que le dé un ejemplo resuelto y después otro sin resolver. No pida la respuesta. La diferencia no está en la herramienta, sino en cuál de los dos trabaja.

Desconfíe de la sensación de entender. Leer una solución clara produce la misma sensación que saber hacerla. Las dos no se distinguen desde dentro, solo con el papel en blanco.

Si con la IA tarda mucho menos, desconfíe. La prueba de la semana siguiente le dirá si lo que ahorró fue tiempo o aprendizaje.

Busque ante quién responder. Un compañero con el que quedar, una academia, un preparador, una fecha de examen. Pagar a alguien para que espere su trabajo no es una debilidad: es lo que hace con sus hijos casi la mitad de las familias. Quien no tiene esa figura y cree no necesitarla pertenece al 20 % o se engaña, y la prueba de la semana siguiente le dirá cuál de las dos cosas.

Cuide la lectura larga. Es lo que más ha caído y lo que un resumen automático sustituye con más facilidad. Un libro al mes, en papel, sin teléfono en la mesa. Parece un consejo de otra época y es la destreza que PISA señala como la más necesaria precisamente ahora: evaluar, relacionar fuentes, sostener la atención.

10.3 Si tiene hijos en edad escolar

Interésese por el día a día, no por el boletín. En PISA, lo que explica la relación entre apoyo familiar y rendimiento es, sobre todo, el interés de los padres por lo que el hijo aprende cada día. El apoyo que los alumnos dicen recibir ha caído en casi todos los países desde 2022.¹³

Mire el tiempo. Una tarea difícil terminada en un cuarto de hora y sin un tachón no es una buena noticia: con IA, la nota de los deberes deja de predecir el examen, y el tiempo dedicado no.¹ Pregunte cómo se hace uno de los ejercicios. No hace falta saber la materia para ver si el chico sabe explicarla.

No prohíba lo que no puede vigilar; cambie lo que sí controla. El lugar donde se hacen los deberes, la hora, el teléfono fuera del cuarto, y que alguien pregunte después. En primaria, sin asistentes abiertos: es el criterio chino, y es razonable.

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. Una academia que entrega ejercicios resueltos vende lo que ya regala el teléfono.

10.4 Si enseña

Califique lo que ha visto hacer. Pruebas cortas y frecuentes, en papel, sobre lo trabajado en casa. Preguntas orales. Que el trabajo largo se defienda de palabra. Los deberes siguen sirviendo para practicar; ya no sirven para poner nota.

Primero a mano. En cada destreza, la herramienta entra después de que el alumno haya demostrado que sabe hacerlo sin ella.

Use la IA para usted. Preparar materiales, variantes de ejercicios, adaptaciones para quien va por detrás. Es donde hay tiempo que ganar y ningún aprendizaje que perder.

Enseñe a dudar de la máquina con casos concretos. Una respuesta generada con un error, y que lo encuentren. PISA lo asocia a mejores resultados entre quienes más la usan.

No base la evaluación en detectar el texto generado. Rediseñe la tarea para que no compense generarlo.

10.5 Si decide sobre un centro o sobre un presupuesto

Tutoría en grupo pequeño dentro del horario, antes que bajar ratios para todos. Es lo que tiene evidencia, y donde una IA de apoyo al tutor es barata y tiene efectos medidos.^{33,37,50}

Ningún despliegue sin grupo de comparación. Un piloto con sorteo y un examen sin la herramienta cuesta poco y evita gastarse lo que se gastó Corea.

Conserve y publique las evaluaciones externas. Son la única medida que la IA no ha estropeado. Sin ellas, las notas internas subirán y nadie sabrá lo que pasa.

Tiempo para el profesor. Rehacer la evaluación cuesta horas de corrección, y el docente español ya declara más carga administrativa que la media.

10.6 Lo que no sirve

Esperar al asistente perfecto. Prohibir el chatbot en el centro y dejar iguales los deberes. Los cursos de «uso responsable» sin ningún cambio en lo que cuenta para la nota. Anunciar resultados antes de medirlos. Y el tremendismo: no es la generación más ignorante de la historia; rinde peor que la de hace quince años en varias pruebas, y eso tiene arreglo.

10.7 Lo que queda por encima de todo esto

Una persona que ha aprendido algo difícil sabe una cosa de sí misma que no se obtiene de otro modo: que puede. La escuela, además de enseñar contenidos, es el sitio donde la mayoría lo descubre por primera vez, 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, el compañero con el que se queda a estudiar no son sustitutos imperfectos de un tutor automático. Son la parte que nunca fue un problema de información.

En buena parte del mundo, esa persona no existe: de aquí a 2030 harán falta 44 millones de profesores más.²³ Allí, un asistente que responde por el teléfono puede ser lo primero que se parece a un maestro, y merece más inversión y más ensayos que todos los debates europeos sobre si se permite en clase. Las dos cosas son verdad a la vez, y ninguna de las dos es el fin de la educación ni su salvación.

Síntesis de veredictos

JUICIOS PRINCIPALES

- 1. La IA ha vuelto gratuita la explicación, que nunca fue lo escaso, y no produce lo que falta: un motivo para sentarse.** Es la versión escolar de la idea que recorre los tres informes: lo que hay que proteger es saber hacerlo sin la máquina.
- 2. Usar IA para los deberes sube la nota de los deberes y baja la del examen.** En el mejor estudio disponible, un 18 % más en los primeros y un 20 % menos en el segundo a los seis meses, y hasta un 24 % menos en los exámenes de acceso a los dos años.
- 3. No daña la herramienta, sino el uso que elige la mayoría.** Cuatro de cada cinco alumnos la usan para acabar antes. El quinto, que mantiene el tiempo de trabajo, aprende lo mismo que sin ella.
- 4. El primer daño es a la medida.** La nota de lo hecho en casa ha dejado de decir lo que el alumno sabe, y la pérdida solo se ve en un examen sin máquina: a los seis meses en el mensual, y creciendo unos dos años en los de acceso.
- 5. PISA 2025 es el peor resultado de la serie y la IA no lo originó.** La lectura cae desde 2012 y las matemáticas desde 2018. La OCDE no nombra la IA al explicar la caída. Lo que retrata es menos atención sostenida y menos ganas. Cuánto añade la IA, PISA no lo puede medir; la hipótesis de este informe es que actúa como acelerante en un punto: los deberes, cuyo tiempo ha caído un 11 % en tres años.
- 6. Peor que hace quince años, no que hace cien.** En 1910 no sabía leer y escribir la mitad de los españoles. Se ha perdido una parte de lo ganado, y fuera de los países ricos el problema

es otro: que no hay profesor.

7. **El tutor que no da la respuesta existe y casi nadie lo usa a fondo.** En los cuatro programas con datos, el uso real va de una vigésima parte a un tercio, según la medida: son problemas de adopción y de continuidad, no una medida directa del esfuerzo.
8. **Lo escaso es la obligación, y no es pereza moral: es sesgo hacia el presente.** Los plazos que impone otro rinden más que los que uno se pone. Casa con que casi la mitad de las familias españolas siga pagando clases particulares, 2.782 millones al año, cuando la explicación ya es gratis.
9. **La mejora con IA vive en la medida cercana.** Los ensayos serios dan efectos entre nulos y modestos cuando se examina sin la herramienta y pasado un tiempo; la excepción es Nigeria, con adulto y horario. El metaanálisis más citado a favor ha sido retractado.
10. **Donde un tutor de IA ha funcionado con alumnos corrientes había un sitio, una hora, un adulto y una prueba sin máquina.** Y el uso con mejor evidencia es la IA detrás del adulto, no delante del alumno.
11. **Prohibir no es la palanca.** Prohibir el móvil mueve las notas entre nada y un poco; España lo ha hecho y ha caído igual. El daño medido de la IA ocurre en casa.
12. **España apuesta 2.800 millones al año a una medida sin efecto medio demostrado,** la bajada de ratios, y no a la que lo tiene, la tutoría en horario escolar.
13. **Lo que hay que cambiar es la evaluación:** primero a mano, examen corto y frecuente sin máquina, mirar el tiempo y no la nota, y un adulto que espere el trabajo.

Tabla 3 Veredictos del informe. Probabilidad de aquí a 2031

Escenario	Probabili- dad	Efecto si ocurre
Las tareas hechas en casa dejan de servir como medida de lo que sabe un alumno	Muy proba- ble	Grave; ya ocurre
La mayoría de los alumnos con acceso libre a la IA la usa para acabar antes	Muy proba- ble	Grave y silencioso
Tutores de IA de buena calidad, gratuitos o casi, al alcance de cualquiera	Muy proba- ble	Beneficioso en potencia
La mayoría del profesorado usa IA para preparar clases	Muy proba- ble	Beneficioso
Nuevos anuncios de resultados espectaculares sin grupo de comparación	Muy proba- ble	Limitado; desorien- ta
PISA 2029 igual o peor que 2025 en la media de la OCDE	Probable	Grave
Suben las notas internas mientras bajan las de las pruebas externas	Probable	Grave; oculta el problema
La evaluación se desplaza hacia pruebas presenciales, orales y en papel	Probable	Beneficioso
Aprobación de la ley de ratios sin mejora apreciable en evaluaciones externas	Probable	Grave para el pre- supuesto
Efectos grandes de tutores de IA con adulto presente donde faltan profesores	Posible	Muy beneficioso
IA como copiloto de tutores y profesores noveles, con efectos medidos	Posible	Beneficioso y bara- to
Un programa nacional de IA educativa con mejora demostrada en un examen externo	Improbable	Beneficioso
Un tutor de IA voluntario mejora de forma apreciable al alumno medio de un país rico	Improbable	—
La IA como causa principal de la caída medida en PISA 2025	Muy impro- bable	—

continúa

Escenario	Probabili- dad	Efecto si ocurre
Una generación «incapaz de pensar» por culpa de la IA	Muy impro- bable	Catastrófico
Sustitución de profesores por tutores de IA en la enseñanza obligatoria de un país rico	Muy impro- bable	Grave

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, a diferencia de los otros dos de la serie, varios escenarios son favorables: la última columna dice entonces qué se ganaría.



Muy improbable
menos del 5 %



Improbable
5–20 %



Posible
20–50 %



Probable
50–80 %



Muy probable
más del 80 %

Referencias

Numeradas por orden de primera cita. Bajo cada entrada, el alcance del resultado que se utiliza. 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 ↗
CEPR Discussion Paper 21577; columna de los autores en VoxEU, 4 de septiembre de 2026. Documento de trabajo.
- 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 ↗
Preprint sin revisión por pares; parte de los autores trabaja en McGraw Hill, dueña de la plataforma.
- 3 Bastani et al.. *Generative AI without guardrails can harm learning*. 25-06-2025. doi.org ↗
Experimento escolar de matemáticas en Turquía; distingue rendimiento con ayuda de aprendizaje posterior. No prueba daño universal ni irreversible.
- 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 ↗
Columna de los autores en VoxEU.
- 6 D. T. Campbell. *Assessing the impact of planned social change*. 1979. doi.org ↗
Evaluation and Program Planning, 2(1). Cuanto más se usa un indicador para decidir, más se corrompe y más corrompe lo que mide.
- 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 ↗
PNAS, 116(39).
- 8 N. C. Soderstrom y R. A. Bjork. *Learning Versus Performance: An Integrative Review*. 2015. doi.org ↗

Perspectives on Psychological Science, 10(2). El rendimiento durante la práctica no indica lo que se retendrá; las «dificultades deseables» lo empeoran a corto plazo y lo mejoran a largo.

- 9 Lira, B., Rogers, T., Goldstein, D. G., Ungar, L. y Duckworth, A. L.. *AI assistance can enhance rather than hinder skill development*. 2025-02. arxiv.org ↗
Ensayos aleatorizados prerregistrados con adultos.
- 10 OCDE. *PISA 2025 Results (Volume I): Future-Ready Students. Executive summary y tabla I.1*. 2026-09-08. oecd.org ↗
- 11 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 ↗
Nota de prensa oficial con las medias de España.
- 12 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 ↗
- 13 OCDE. *PISA 2025 Results (Volume I), capítulo 4: Student school life and beyond*. 2026-09-08. oecd.org ↗
Dispositivos, IA, deberes, familia y absentismo.
- 14 OCDE. *PISA 2025 Results (Volume I)*. 2026-09-08. oecd.org ↗
760.000 alumnos de 15 años en 91 países. Las relaciones entre uso de IA y puntuación son correlaciones.
- 15 OCDE. *PISA 2025 Results (Volume I). Country notes: Spain*. 2026-09-08. oecd.org ↗
- 16 Infobae. *Asturias, la Comunidad de Madrid y Castilla y León despuntan mientras que la Comunidad Valenciana y País Vasco se hunden en el informe PISA 2025*. 2026-09-08. infobae.com ↗
Resultados por comunidades, de prensa.
- 17 De Gabriel, N.. *Alfabetización, semialfabetización y analfabetismo en España (1860-1991)*. 1997. revistas.ucm.es ↗
Revista Complutense de Educación, 8(1).
- 18 National Center for Education Statistics. *NAEP Long-Term Trend Assessment Results: Reading and Mathematics, age 13*. 2023-06. nationsreportcard.gov ↗
- 19 UNESCO. *Countries commit to reducing the number of out-of-school children by 165 million*. 2025. unesco.org ↗

Estimación de 272 millones de niños y jóvenes sin escolarizar.

- 20 Bratsberg, B. y Rogeberg, O.. *Flynn effect and its reversal are both environmentally caused*. 2018. pnas.org ↗
PNAS, 115(26).
- 21 National Center for Education Statistics. *NAEP Long-Term Trend Assessment Results 2025*. 2026. nationsreportcard.gov ↗
Los alumnos de 9 años suben cuatro puntos en lectura y en matemáticas respecto a 2022; en los de 13 años no hay cambios significativos respecto a 2023.
- 22 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 ↗
- 23 UNESCO y Equipo Especial Internacional sobre Docentes. *Global Report on Teachers*. 2024. unesco.org ↗
- 24 Banco Mundial. *World Development Report 2026: The Promise of Artificial Intelligence*. 2026-08-04. worldbank.org ↗
- 25 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 ↗
Blog del Banco Mundial; resultados de la evaluación aleatorizada del piloto del estado de Edo.
- 26 Oreopoulos, P. y Low, N.. *One Click Away: AI Tutoring with Khanmigo in a Two-Year School Experiment*. 2026-08. nber.org ↗
NBER Working Paper 35620.
- 27 Moreno, J. M. y Gortazar, L.. *El mercado de las clases particulares en España entre 2020 y 2024*. 2026-06-25. esade.edu ↗
EsadeEcPol, con la Encuesta de Gasto de los Hogares en Educación del INE.
- 28 Chalkbeat. *Students rarely engaged with Khan Academy's AI-powered tutor Khanmigo, study finds*. 2026-08-25. chalkbeat.org ↗
Estudio de dos años en distritos de Estados Unidos.
- 29 Holt, L.. *The 5 Percent Problem: Online mathematics programs may benefit most the kids who need it least*. 2024. educationnext.org ↗
Education Next, 24(4).
- 30 TI-Hüpe (AI Leap Foundation). *Estonia Launches Nationwide AI Leap Education Program*. 2026. tihupe.ee ↗
Cifras de activación comunicadas por el propio programa, marzo de 2026.

- 31 D. Ariely y K. Wertenbroch. *Procrastination, Deadlines, and Performance: Self-Control by Precommitment*. 2002. [doi.org](#) ↗
Psychological Science, 13(3). Los plazos impuestos desde fuera y repartidos rinden más que los que uno se pone.
- 32 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](#) ↗
El artículo original, de J. Wang y W. Fan, se publicó el 6 de mayo de 2025.
- 33 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](#) ↗
NBER Working Paper 27476; publicado en American Educational Research Journal en 2024.
- 34 B. S. Bloom. *The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring*. 1984. [doi.org](#) ↗
Educational Researcher, 13(6).
- 35 Kestin et al. (Harvard). *AI tutoring outperforms in-class active learning*. 2025-06. [nature.com](#) ↗
194 universitarios de Física en Harvard: alumnos muy seleccionados y motivados.
- 36 Google DeepMind y Eedi. *AI tutoring can safely and effectively support students: An exploratory RCT in UK classrooms*. 2025-12. [arxiv.org](#) ↗
165 alumnos de 13 a 15 años; los mensajes de la IA pasaban por un tutor humano.
- 37 Wang et al. (Stanford). *Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise*. 2024-10. [nssa.stanford.edu](#) ↗
Ensayo aleatorizado con 900 tutores y 1.800 alumnos de 3.º a 8.º, durante dos meses. No detecta mejora estadísticamente significativa en el examen de fin de curso, con una exposición breve y escasa.
- 38 OpenAI. *New tools for understanding AI and learning outcomes*. 2026-03-04. [openai.com](#) ↗
Ensayo propio de la empresa sobre su «modo estudio».
- 39 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](#) ↗
- 40 La Noticia SV. *PISA 2025: El Salvador mejora sus resultados en Ciencias, Matemáticas y Lectura*. 2026-09. [lanoticiasv.com](#) ↗
Cifras nacionales de PISA 2025, de prensa.

- 41 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 ↗
NBER Working Paper 35132, revisado en septiembre de 2026.
- 42 Lynnwood Times. *Scores from El Salvador Students in AI Tutoring Pilot Matches Germany and Sweden on PISA Test*. 2026-09-06. lynnwoodtimes.com ↗
1.198 alumnos voluntarios de 171 escuelas ya incluidas en un programa de mejora; sin grupo de control ni medida previa.
- 43 Rest of World. *South Korea's AI textbooks fail after rushed rollout*. 2025. restofworld.org ↗
Reportaje; la adopción cayó del 37 al 19 % y el Parlamento retiró a los libros su rango oficial.
- 44 CNBC. *Key AI hub China restricts schoolchildren's use of the tech*. 2025-05-15. cnbc.com ↗
Guías del Ministerio de Educación chino de 12 de mayo de 2025.
- 45 Comité de orientación de la enseñanza básica del Ministerio de Educación de China. *Guía de uso de la inteligencia artificial generativa en la enseñanza primaria y secundaria (edición de 2025)*. 2025-05-13. hnzj.ttc dw.cn ↗
Texto de la guía, en chino. El apartado IV incluye un punto dirigido a los alumnos (no copiar en los deberes contenido generado) y otro a las familias (supervisión y coordinación con la escuela). No acredita cumplimiento.
- 46 Figlio, D. N. y Özek, U.. *The Impact of Cellphone Bans in Schools on Student Outcomes: Evidence from Florida*. 2025-10. nber.org ↗
NBER Working Paper 34388.
- 47 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 ↗
NBER Working Paper 35233.
- 48 Instituto Nacional de Evaluación Educativa. *Resultados de TALIS 2024: informe español*. 2025-10-07. inee.educacion.es ↗
- 49 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 ↗
Journal for Research in Mathematics Education, 34(5).
- 50 Gortazar, L. y Montalbán, J.. *La nueva ley de ratios: estimación de su impacto económico*. 2026-05-26. esade.edu ↗

Centro de políticas económicas de Esade (EsadeEcPol).

- 51 Roediger, H. L. y Karpicke, J. D.. *Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention*. 2006. [doi.org](#) ↗
Psychological Science, 17(3).