

Diversidad de muestras, diversidad de resultados: lecciones internacionales desde PISA

1. Introducción

Desde su origen, las evaluaciones internacionales comparadas han estado en el centro del escrutinio académico y político, ocupando un lugar privilegiado en la agenda educativa global. Entre ellas, el Programa para la Evaluación Internacional de Estudiantes (PISA, por sus siglas en inglés), coordinado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), se ha consolidado desde el año 2000 como uno de los instrumentos más influyentes para medir las competencias de los estudiantes de 15 años en lectura, matemáticas y ciencias (Reiss et al., 2019). La periodicidad trienal de sus aplicaciones permite no solo establecer comparaciones entre más de 80 países, sino también identificar tendencias y transformaciones en los sistemas educativos a lo largo del tiempo, dotando a PISA de un prestigio y legitimidad que trascienden el ámbito estrictamente académico para incidir de manera directa en el diseño de políticas públicas, la opinión mediática y la percepción social sobre la “calidad” de la educación.

No obstante, el mismo prestigio que rodea a PISA ha estado acompañado de cuestionamientos recurrentes en torno a sus fundamentos metodológicos. En particular, la selección de las muestras constituye uno de los aspectos más debatidos en la literatura especializada (OCDE, 2017; Meinck, 2020; Skinner, 2014). A diferencia del muestreo aleatorio simple, PISA utiliza un diseño complejo que combina probabilidad proporcional al tamaño (PPS) y procedimientos de estratificación explícita e implícita, con el fin de aumentar la eficiencia de los estimadores y garantizar la representatividad. Esto significa que la selección de escuelas y estudiantes no es completamente al azar, sino que se organiza de manera que refleje mejor la diversidad del sistema educativo de cada país y permita comparar resultados entre ellos con mayor precisión. Estos métodos buscan compensar el hecho de que los estudiantes de una misma escuela tienden a ser más

similares entre sí que los de diferentes centros, lo que hace necesario recurrir a modelos multinivel para un análisis más riguroso de los resultados (Cai, 2013).

Sin embargo, la sofisticación estadística no ha eliminado las críticas: investigadores han señalado la falta de transparencia en los criterios de exclusión (por ejemplo, no siempre se especifica claramente cuántos estudiantes fueron excluidos por discapacidad, por idioma o por pertenecer a escuelas de difícil acceso), las limitaciones derivadas del tamaño de las muestras en determinados contextos (por ejemplo, en países pequeños o con alta dispersión geográfica, el número de escuelas seleccionadas puede ser insuficiente para representar adecuadamente la diversidad del sistema educativo) y los sesgos introducidos por factores estructurales como la repetición escolar (los niños pueden estar en cualquier grado desde que tengan 15 años), la estratificación social o la baja cobertura educativa en países en desarrollo.

El debate no es menor, ya que las inferencias realizadas a partir de PISA poseen un enorme peso simbólico y político. Rankings internacionales, informes de competitividad educativa y comparaciones de desempeño nacional se construyen sobre la base de estos resultados, influyendo en la dirección de las reformas educativas y en la legitimidad de los sistemas escolares. Así, problemas técnicos aparentemente “marginales”, como la definición de la población objetivo, la tasa de exclusión o la precisión de los estimadores, pueden derivar en sesgos significativos que cuestionan la validez de las comparaciones internacionales (Aßmann et al., 2011; Valliant et al., 2018).

Este informe tiene como propósito analizar la manera en que se construye el marco de evaluación y el diseño muestral de las pruebas PISA, así como sus principales limitaciones, a partir de la revisión de estudios empíricos y aportes teóricos de diversos investigadores en distintos países y de las implicaciones que estos hallazgos tienen para la validez de los resultados. Desde un enfoque pedagógico, se abordarán aspectos como la representatividad de las muestras en contextos nacionales diversos, la influencia de las decisiones técnicas en la construcción de indicadores comparativos y las alternativas metodológicas que podrían contribuir a fortalecer la robustez y legitimidad del programa. En este sentido, el informe busca aportar insumos a un debate de creciente relevancia en el campo de la evaluación educativa internacional.

2. Contexto general de las pruebas PISA

Las pruebas PISA constituyen desde el año 2000 la evaluación educativa comparada de mayor alcance a nivel mundial. Su aplicación trienal a estudiantes de 15 años — independientemente del grado escolar en que se encuentren— busca medir competencias fundamentales en lectura, matemáticas y ciencias, junto con un conjunto de habilidades transversales vinculadas al aprendizaje a lo largo de la vida. La relevancia de esta evaluación no solo radica en su capacidad para estimar el desempeño académico, sino también en su potencial para generar evidencia que orienta políticas públicas, establece comparaciones internacionales y visibiliza desigualdades educativas entre y dentro de los países participantes.

PISA no se limita a reportar promedios de desempeño. Su diseño metodológico parte de la estimación de una distribución poblacional multivariable latente, basada en las respuestas de los estudiantes tanto a los ítems de la prueba como a los cuestionarios contextuales. Para ello, se utiliza el marco de la Teoría de Respuesta al Ítem (TRI), un enfoque psicométrico que modela la probabilidad de que un estudiante responda correctamente a una pregunta en función de su nivel de competencia y de las características del ítem (como su dificultad o capacidad de discriminación). Este método permite calcular indicadores como promedios nacionales, brechas de rendimiento entre subpoblaciones (por ejemplo, según género o nivel socioeconómico) y tendencias a lo largo del tiempo. No obstante, toda estimación derivada de la prueba está sujeta a distintos tipos de incertidumbre estadística: (1) el error de muestreo, asociado al hecho de trabajar con una muestra y no con la población completa; (2) el error de medición, relacionado con las limitaciones de precisión de los instrumentos; y (3) el error de enlace, derivado de la necesidad de calibrar parámetros de ítems entre diferentes rondas de evaluación. Entre estos, el error de muestreo constituye la fuente principal de incertidumbre, lo que otorga un papel central al diseño muestral en la solidez y validez de los resultados obtenidos (OCDE, 2017).

La selección de la muestra en PISA responde a un diseño estratificado y multietápico característico de las evaluaciones internacionales a gran escala. En la primera etapa se seleccionan escuelas utilizando el método de probabilidad proporcional al tamaño (PPS), lo que otorga mayor probabilidad de inclusión a instituciones con mayor número de estudiantes (Meinck, 2020; Skinner, 2014). Posteriormente, en cada escuela seleccionada se elige aleatoriamente entre 30 y 40 estudiantes de 15 años, usualmente tras clasificarlos por variables como género y curso. Este procedimiento corresponde a un muestreo por conglomerados, donde el centro educativo constituye la unidad primaria de muestreo y el estudiante la unidad secundaria.

Un aspecto clave del diseño muestral de PISA es la estratificación, que puede ser explícita o implícita. La estratificación explícita implica dividir a todas las escuelas elegibles en subgrupos homogéneos según ciertas características (p. ej., tipo de gestión, región, nivel socioeconómico), dentro de los cuales se realiza la selección. Por su parte, la estratificación implícita ordena las escuelas en función de atributos específicos y luego aplica un muestreo sistemático, lo que asegura que la muestra conserve aproximadamente las proporciones poblacionales de los estratos. En ambos casos, la lógica es mejorar la eficiencia del diseño y reducir los errores de muestreo: cuando los estratos son internamente homogéneos y entre sí heterogéneos, las estimaciones resultan más precisas y reflejan mejor la diversidad de la población estudiantil (Cochran, 1977; Aßmann et al., 2011; Meinck & Vandenplas, 2021).

Dado que las probabilidades de selección no son iguales para todas las unidades, PISA incorpora ponderaciones muestrales que corrigen estas desigualdades y compensan además la falta de respuesta. Las ponderaciones se calculan combinando las probabilidades inversas de selección en cada etapa con ajustes adicionales, de modo

que los estimadores derivados puedan considerarse insesgados y consistentes bajo el enfoque de Horvitz-Thompson (OCDE, 2017). Para el cálculo de varianzas, se utiliza la técnica de Replicación Repetida Balanceada (BRR) con ajuste de Fay, que permite tener en cuenta la estructura compleja del diseño. Este procedimiento se implementa a partir de la creación de “zonas de varianza” conformadas por pares de escuelas muestreadas con características similares, sobre las cuales se construyen réplicas que garantizan la validez de las estimaciones de error (Rust & Rao, 1996; Judkins, 1990).

En síntesis, PISA descansa sobre un diseño muestral técnicamente sofisticado que combina muestreo estratificado, selección por conglomerados, ponderaciones y métodos avanzados de estimación. Sin embargo, esta complejidad metodológica no ha estado exenta de críticas. La exclusión de ciertos grupos de estudiantes, las diferencias en la implementación nacional del muestreo y la sensibilidad de los resultados a las decisiones técnicas de estratificación han suscitado interrogantes sobre la validez y comparabilidad de los puntajes reportados. De allí la importancia de examinar con detenimiento cómo los distintos países adaptan y ejecutan el diseño muestral de PISA, pues estas variaciones pueden incidir de manera significativa en las conclusiones que se derivan de la prueba y, en última instancia, en las políticas educativas que se sustentan en ella.

3. Importancia del diseño muestral en evaluaciones internacionales

En evaluaciones internacionales a gran escala como PISA, el diseño muestral no constituye un aspecto meramente técnico o accesorio, sino un componente central que determina la validez de las inferencias y la credibilidad de los hallazgos. El principio rector es claro: una muestra representativa permite que los resultados obtenidos se generalicen a la población objetivo con un nivel de confianza estadístico aceptable; por el contrario, una muestra sesgada o incompleta limita gravemente la posibilidad de realizar comparaciones y de extraer conclusiones útiles para la política educativa.

La estadística de inferencia se apoya en el supuesto de aleatoriedad. El muestreo aleatorio permite que cada unidad de la población tenga una probabilidad conocida y distinta de cero de ser seleccionada, lo que garantiza que las estimaciones no estén contaminadas por sesgos sistemáticos. Por el contrario, los problemas derivados de muestras no aleatorias son bien conocidos en la tradición científica. El ejemplo clásico del Informe Kinsey (1948) sobre la sexualidad en EE. UU. muestra cómo una estrategia de muestreo de bola de nieve condujo a conclusiones cuestionadas: una proporción elevada de los participantes provenía de poblaciones atípicas (como hombres encarcelados), lo que distorsionó la imagen del conjunto. John Tukey sintetizó la lección metodológica con una frase célebre: “una selección aleatoria de tres personas habría sido mejor que un grupo de 300 elegido de manera sesgada” (Leonhardt, 2000). Casos recientes como la crisis de replicación en psicología (Open Science Collaboration, 2015) confirman que la falta de muestras representativas, sumada a otros sesgos como el de publicación, contribuye a la fragilidad de la evidencia científica.

A diferencia de estos estudios con muestras problemáticas, las evaluaciones internacionales suelen considerarse de alta calidad metodológica. Su diseño multietápico y estratificado, sumado a la inclusión de miles de estudiantes por país, reduce el margen de error y ofrece estimaciones con alta precisión. Desde la perspectiva estricta del muestreo, PISA proporciona datos comparables entre países con un rigor poco frecuente en la investigación educativa. Sin embargo, esta afirmación debe matizarse: la calidad de un diseño muestral no se deriva únicamente de su sofisticación teórica, sino de su implementación práctica en contextos heterogéneos.

En primer lugar, la exclusión de estudiantes constituye un área crítica. PISA establece que hasta un 5 % de la población objetivo puede ser excluida, por ejemplo, en el caso de alumnos que no dominan la lengua de la prueba o que presentan necesidades educativas especiales. No obstante, algunos países han sobrepasado este umbral. El caso de Suecia en PISA 2018, que excluyó más del 10 % de su cohorte, provocó un intenso debate público y académico (Andersson & Sandgren Massih, 2023). Si bien las exclusiones aleatorias no comprometen necesariamente la representatividad, investigaciones muestran que las no respuestas o exclusiones tienden a correlacionarse con el estatus socioeconómico y con el rendimiento de los estudiantes (Anders et al., 2020; Durrant & Schnepf, 2018; Micklewright et al., 2012; Jerrim, 2021). Esto significa que los grupos más vulnerables suelen estar subrepresentados, lo cual sesga las estimaciones hacia una visión demasiado optimista del desempeño de un país.

En segundo lugar, la definición de la población objetivo plantea desafíos conceptuales y políticos. PISA evalúa únicamente a jóvenes de 15 años matriculados en instituciones educativas formales. En países del Norte Global, donde la cobertura educativa y el porcentaje de los jóvenes que llegan “a tiempo” a los 15 años se acerca al 100 %, esta definición es poco problemática. Sin embargo, en contextos con tasas de escolarización más bajas o con trayectorias educativas más irregulares, el panorama cambia de forma sustancial. Por ejemplo, en Argentina el 81 % de los chicos de 15 años realizan su escolaridad sin repetir, abandonar o entrar tarde al sistema educativo. En Chile y Brasil esta cifra es del 79 %, seguidos de Perú (77 %), México (69 %), Paraguay (69 %) y Uruguay (59 %). Los niveles más altos de sobreedad se encuentran en Colombia, donde solo el 53 % de los estudiantes de 15 años están en el grado esperado (OCDE, 2023). Así, los resultados de PISA en estos países representan condiciones muy diversas de estudiantes de 15 años, quienes, debido a trayectorias escolares irregulares, pueden encontrarse en grados distintos al momento de la evaluación, lo que plantea dudas sobre la pertinencia de la población objetivo como referente universal.

Por último, las exclusiones sistemáticas de estudiantes con necesidades educativas especiales reflejan otra tensión entre criterios de comparabilidad y principios de equidad. Aunque algunos países han avanzado hacia una mayor inclusión, las diferencias entre naciones son notorias y aún poco comprendidas. Persistir en este patrón implica que PISA, en la práctica, sigue evaluando un ideal implícito de “estudiante típico”, lo que limita su capacidad para reflejar la diversidad real de las poblaciones escolares.

En suma, el diseño muestral en evaluaciones como PISA cumple una doble función: es la condición de posibilidad para producir estimaciones estadísticamente sólidas y, al mismo tiempo, el punto de mayor vulnerabilidad de los resultados. Cualquier desviación en su aplicación —ya sea por exclusiones excesivas, por definiciones restrictivas de la población o por fallas en el control de la no respuesta— no solo afecta la precisión técnica de las estimaciones, sino también la interpretación política y social de los resultados. Esto es crucial, dado que los datos de PISA se utilizan para orientar reformas educativas, distribuir recursos y elaborar rankings internacionales que inciden en la reputación de los sistemas escolares. Así, lo que a primera vista puede parecer un asunto estadístico tiene, en realidad, implicaciones de gran alcance para la equidad educativa y para la legitimidad de las políticas públicas que se apoyan en esta evidencia.

4. Principales críticas al muestreo de PISA

Diversos estudios han puesto de relieve las limitaciones del diseño muestral de PISA y la manera en que estas pueden incidir en la validez de los resultados. Fernández-Cano (2016) cuestiona la transparencia del proceso, al señalar que la OCDE no divulga con suficiente claridad los criterios técnicos empleados en la selección y sustitución de escuelas o estudiantes. Esta opacidad metodológica dificulta la replicabilidad de los análisis externos y debilita la confianza en la neutralidad del estudio, más aún cuando sus resultados tienen un fuerte impacto en la formulación de políticas educativas y en la reputación internacional de los sistemas escolares evaluados.

A estas observaciones se suman críticas sobre la exclusión sistemática de ciertos grupos de estudiantes, como aquellos con necesidades educativas especiales, ausencias prolongadas o trayectorias escolares atípicas. Krohne, Meier y Tillmann (2004) advierten que este sesgo de selección tiende a favorecer a estudiantes con mejores expectativas de rendimiento, inflando artificialmente los resultados nacionales. Investigaciones posteriores, como la de Spaul (2018), muestran que en contextos de baja escolarización —caso de Turquía— PISA termina evaluando únicamente a una fracción de la cohorte de 15 años, lo que convierte la muestra en un reflejo de una élite escolar y no de toda la población juvenil. Estas dinámicas generan serias dudas sobre la comparabilidad entre países con realidades educativas muy dispares.

Otros investigadores han identificado problemas de diseño más específicos que también afectan la representatividad y precisión de los resultados. Mortimore (2009) advierte sobre la posibilidad de manipulaciones estratégicas en la conformación de la muestra, mientras que Mang (2023) documenta cómo en sistemas federales, como el alemán, la estratificación por regiones puede producir estratos vacíos y reducir la eficiencia estadística. Asimismo, Caro y Kyriakides (2019) destacan el sesgo introducido por las fechas de corte escolar, que generan cohortes heterogéneas en términos de tiempo de escolarización. Aunque PISA utiliza ponderaciones para corregir estas distorsiones (Fuchs & Woessmann, 2004), las críticas coinciden en que las limitaciones del diseño

muestral condicionan la interpretación de los resultados y obligan a analizarlos con cautela.

En síntesis, las críticas coinciden en que los procedimientos de muestreo de PISA, lejos de ser meramente técnicos, tienen implicaciones directas sobre la validez, representatividad y comparabilidad de los resultados. Estas limitaciones se expresan en múltiples dimensiones: desde la falta de transparencia en los criterios aplicados, pasando por la exclusión sistemática de ciertos grupos de estudiantes, hasta los sesgos introducidos por contextos de baja escolarización o por decisiones de diseño como las fechas de corte y la estratificación territorial. Con el fin de analizar estos cuestionamientos de manera estructurada, a continuación, se presentan las principales críticas al muestreo de PISA organizadas en torno a sus diferentes ejes problemáticos.

Falta de transparencia en el muestreo

Una de las críticas más recurrentes al muestreo de PISA es su opacidad metodológica. Aunque la OCDE publica una amplia documentación técnica sobre tasas de exclusión, tamaños muestrales y procedimientos de selección, diversos autores señalan que no siempre resulta claro cómo se aplican estos criterios en la práctica ni qué acuerdos internacionales median en situaciones críticas. Fernández-Cano (2016) describe este proceso como “opaco”, subrayando que la ausencia de plena transparencia dificulta la replicabilidad de los análisis y, en consecuencia, mina la confianza en la neutralidad del proceso. Esta crítica apunta a un problema de fondo: si bien PISA se presenta como una herramienta objetiva de diagnóstico educativo, la falta de claridad metodológica puede dar pie a sospechas sobre la comparabilidad real de los resultados, especialmente cuando estos se utilizan con fines políticos.

Exclusión de estudiantes y representatividad

La exclusión de ciertos grupos de estudiantes es otro de los puntos más controvertidos. De acuerdo con los lineamientos oficiales, PISA permite la exclusión de hasta un 5 % de la población objetivo, lo que suele aplicarse a estudiantes con necesidades educativas especiales, aquellos con dominio insuficiente del idioma de la prueba o jóvenes con retrasos escolares severos. Sin embargo, en la práctica, este umbral ha sido ampliamente superado: en 2006 se excluyó aproximadamente un 15 % de los estudiantes, en 2009 un 25 % y en 2012 un 20 % (Fernández-Cano, 2016). Estas tasas de exclusión, además de variar significativamente entre países, generan un sesgo diferencial que compromete la comparabilidad internacional. Investigadores como Krohne, Meier y Tillmann (2004) ya habían advertido que la exclusión sistemática de ciertos grupos, junto con la ausencia de itinerarios claros para estudiantes repetidores, distorsiona la representatividad de las muestras y, por tanto, limita la validez de las inferencias.

Los efectos de esta práctica son particularmente graves en países donde el acceso a la escolarización a los 15 años no es universal. Spaul (2018), por ejemplo, muestra que, en Turquía, alrededor de un tercio de la población de la edad objetivo no estaba escolarizada en PISA 2003. En consecuencia, los resultados publicados sobrestimaban el desempeño

real del país, al excluir a una franja significativa de jóvenes. Este sesgo no solo afecta la medición del aprendizaje, sino que también condiciona las comparaciones internacionales, favoreciendo a los países con mayor cobertura educativa.

Comparabilidad entre países con contextos distintos

La comparabilidad internacional constituye otro de los grandes dilemas metodológicos de PISA. En países del Norte Global, donde prácticamente todos los jóvenes de 15 años permanecen escolarizados, los resultados pueden interpretarse como representativos de toda la cohorte. En cambio, en contextos de América Latina, África o Asia, una proporción importante de adolescentes se encuentra fuera del sistema educativo, lo que deja fuera de la evaluación a los sectores más vulnerables. Fernández-Cano (2016) subraya que este diseño rígido, que define la población objetivo exclusivamente a partir de la edad y no del grado escolar alcanzado, genera saltos muestrales difíciles de justificar y reduce la pertinencia de las comparaciones entre sistemas tan heterogéneos (Strietholt, Rosén & Bos, 2013).

Además, la imposición del modelo evaluativo de la OCDE plantea una paradoja: si bien la participación en PISA es voluntaria para los ministerios de educación, los estudiantes, familias y docentes rara vez son consultados o considerados en el proceso (Fernández-Cano, 2016). De esta manera, PISA se convierte en una evaluación externa sumativa que responde más a intereses políticos y comparativos que a demandas internas de los sistemas educativos.

Tamaños muestrales y sesgos potenciales

Otra crítica se dirige a los tamaños muestrales y a su aparente falta de proporcionalidad respecto a la población total de cada país. Un caso ilustrativo es el de PISA 2009: Austria, con alrededor de 8 millones de habitantes, incluyó en la muestra a 6.590 estudiantes, mientras que Japón, con una población de 126 millones, evaluó a solo 6.088 (OECD, 2012). A primera vista, esta disparidad podría interpretarse como una inequidad en la representación relativa. Sin embargo, debe considerarse que la lógica del muestreo estadístico no exige que el tamaño de la muestra crezca de manera proporcional a la población total. Lo que prima es alcanzar un número mínimo de casos que garantice inferencias robustas, lo cual explica por qué un país con población mucho mayor no requiere necesariamente una muestra de tamaño equivalente en proporción. Pese a estas justificaciones metodológicas, la percepción de asimetría permanece y alimenta el debate sobre la transparencia, la comparabilidad y la legitimidad de los resultados.

Racionalidad inconsistente e inferencias causales

Más allá de los aspectos técnicos, algunos autores cuestionan la racionalidad subyacente a PISA. Fernández-Cano (2016) sostiene que el modelo evaluativo de la OCDE adolece de inconsistencias y presupuestos simplistas. PISA, por ejemplo, parte de la hipótesis de que la riqueza económica correlaciona positivamente con el rendimiento escolar, y que mejorar ciertas características escolares redundará en una mejora de los resultados y, en última instancia, en el desarrollo económico. Sin embargo, estudios como los de Bank

(2012) y De Witte & Kortelainen (2013) señalan que estas correlaciones son débiles, variables entre países y, sobre todo, incapaces de establecer relaciones causales. Pese a ello, PISA utiliza modelos estadísticos como regresiones múltiples o multinivel para extraer inferencias que a menudo son interpretadas por los medios y los gobiernos como si fueran conclusiones causales. Esta confusión metodológica no solo es problemática en términos técnicos, sino que también tiene implicaciones políticas de gran alcance: políticas educativas nacionales pueden diseñarse en función de correlaciones espurias o sobredimensionadas.

Inclusión de estudiantes con necesidades educativas especiales (NEE)

Por último, la inclusión de estudiantes con NEE constituye un ámbito en el que PISA enfrenta críticas crecientes. LeRoy et al. (2018) advierten que existen fuertes diferencias entre países respecto a los criterios de exclusión e inclusión de estos estudiantes, lo que genera falta de consistencia en la comparabilidad. Proponen dos medidas clave: reducir la variabilidad en las tasas de exclusión e implementar un sistema estandarizado para clasificar las condiciones y necesidades especiales, que incluya categorías físicas, sensoriales, intelectuales, conductuales y de alfabetización lingüística. Además, señalan que la falta de una política clara de adaptaciones en las pruebas limita la participación de este grupo, reforzando la idea de que PISA sigue evaluando a un perfil de “estudiante típico” y no refleja plenamente la diversidad escolar existente.

5. Análisis por país

Colombia

En 2022, Colombia evaluó a 7.804 estudiantes mediante formato digital con dispositivos USB, de los 9.200 inicialmente previstos. La muestra representó a 532.000 estudiantes de 7.º a 11.º grado, nacidos entre enero y diciembre de 2007, provenientes de 264 establecimientos educativos de 28 departamentos. El 21 % asistía a colegios privados y el 79 % a oficiales (40 % urbanos y 39 % rurales). En términos de cobertura, el número de estudiantes evaluados ha crecido en el tiempo: de 4.478 en 2006 a 7.804 en 2022, con un pico de participación en 2015 (11.777). Tras la caída en 2018 (7.522), la cifra volvió a aumentar en 2022, aunque de manera más moderada.

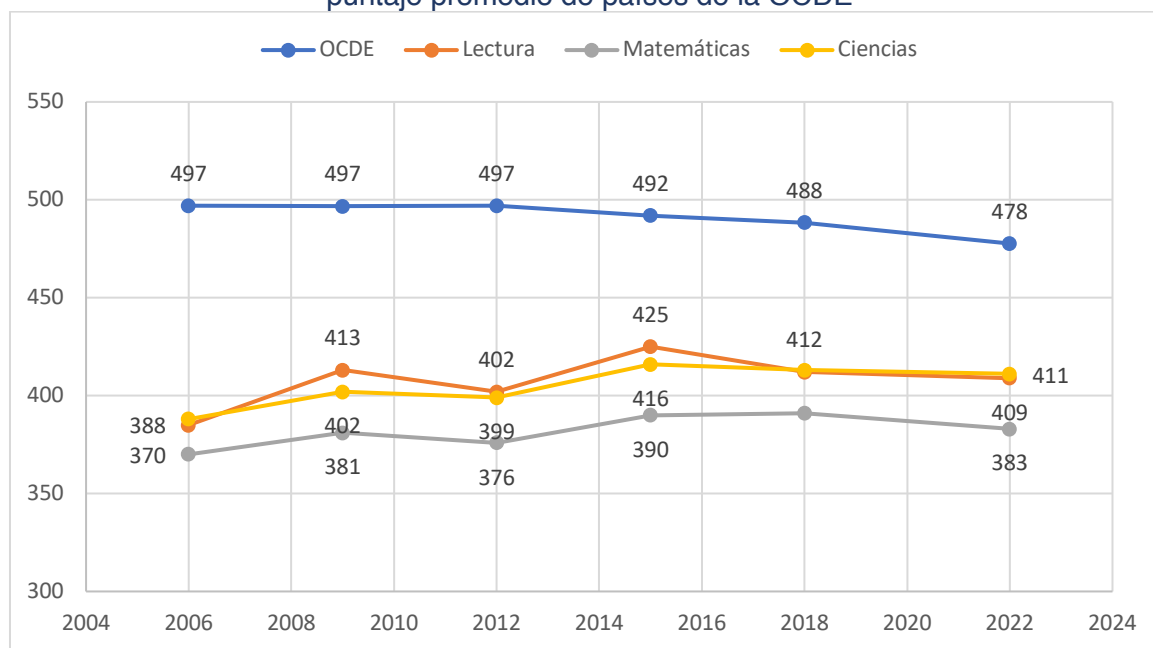
En cuanto a la distribución por sexo, desde 2006 hasta 2015 las mujeres superaron a los hombres entre 4 y 8 puntos porcentuales (p.p.). Aunque esta diferencia se mantuvo en 2018 y 2022, el margen se redujo a solo 2 p.p. en favor de las mujeres. Respecto al nivel socioeconómico, el 27 % de los estudiantes colombianos se ubicó en el primer quintil del índice ESCS¹ (el más bajo), mientras que solo el 12 % alcanzó el quinto quintil, en contraste con el promedio de la OCDE, donde 37 % se ubicó en el quintil superior y apenas 5 % en el más bajo.

¹ Índice de estatus económico, social y cultural (ESCS, por sus siglas en inglés).

La evolución de la cobertura institucional muestra que las escuelas rurales oficiales han aumentado su participación, pasando de 85 en 2015 a 110 en 2022, mientras que los establecimientos privados se redujeron de 88 a 56 en el mismo periodo. En los colegios urbanos oficiales la caída fue aún más marcada, pasando de 134 en 2015 a 84 en 2022. Esto se refleja en la distribución poblacional: la proporción de estudiantes en colegios urbanos oficiales bajó de 48 % en 2015 a 40 % en 2022, mientras que en los rurales subió de 27 % a 39 %. Además, en 2022 se aplicó una sobremuestra en Bogotá (56 colegios y 2.300 estudiantes), que aumentó la participación de la capital. Por último, el 2,9 % de los evaluados eran inmigrantes (primera o segunda generación), cifra inferior al promedio latinoamericano y muy por debajo de la OCDE.

En cuanto a resultados, Colombia obtuvo 383 puntos en matemáticas, 409 en lectura y 411 en ciencias, ubicándose en el sexto lugar de América Latina, detrás de Chile, Uruguay, México, Perú y Costa Rica. Si bien se registraron descensos respecto a 2018, estas caídas fueron menores a las observadas en el promedio OCDE. En lectura, la reducción fue de solo 3 puntos frente a 11 en la OCDE, y en ciencias de 2 puntos frente a 4 en la OCDE. La caída más pronunciada se dio en matemáticas, con un puntaje que descendió a niveles preocupantes en relación con los estándares internacionales.

Gráfica 1. Resultados de Colombia en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

En general, el aumento progresivo de la representación rural, acompañado de la disminución de colegios privados y urbanos oficiales, tiende a visibilizar con mayor fuerza las brechas educativas históricas del país, lo que puede haber incidido en el promedio más bajo frente a países que limitan la participación de poblaciones con mayores desventajas. Por otro lado, la sobremuestra aplicada en Bogotá introduce un contrapeso

relevante. Dado que la capital tiende a obtener resultados más altos que el promedio nacional, su sobrerrepresentación pudo atenuar parcialmente el impacto de la mayor participación rural. Esto genera una tensión metodológica: mientras la inclusión de más estudiantes rurales asegura una fotografía más fiel del sistema educativo colombiano, la sobrerrepresentación de Bogotá puede mejorar artificialmente el promedio al equilibrar los resultados hacia arriba.

Finalmente, aunque los resultados de Colombia se mantuvieron por debajo de la media internacional, la menor caída frente a la OCDE refuerza la idea de resiliencia. Sin embargo, esta resiliencia no puede entenderse sin considerar la composición muestral: la inclusión creciente de estudiantes rurales y de bajos recursos no solo explica la magnitud de las brechas internas, sino también la dificultad de establecer comparaciones directas con países que presentan una muestra más sesgada hacia sectores urbanos y favorecidos.

En conclusión, el caso colombiano muestra cómo las decisiones de muestreo — especialmente el peso relativo de colegios rurales, privados y urbanos— inciden directamente en los resultados reportados. Este diseño, aunque desfavorable en términos comparativos, aporta mayor validez interna al estudio, pues refleja con mayor fidelidad la compleja realidad educativa del país.

Argentina

En 2022, Argentina implementó por primera vez la evaluación digital, alcanzando a 12.111 estudiantes de 457 escuelas, lo que representó el 99 % de los colegios y el 86 % de los estudiantes previstos. Para garantizar la cobertura, el Ministerio de Educación proveyó 8.157 computadoras, además de capacitar a tutores y facilitadores encargados de acompañar a los estudiantes en el uso de la plataforma digital. El diseño muestral fue gestionado por la Secretaría de Evaluación e Información Educativa (SEIE) y supervisado por Westat, asegurando su validez técnica. La selección se realizó en dos etapas: primero, una selección probabilística de escuelas según matrícula y, luego, un muestreo aleatorio de estudiantes de 15 años (máximo 35 por escuela), con dos establecimientos de reemplazo por cada centro.

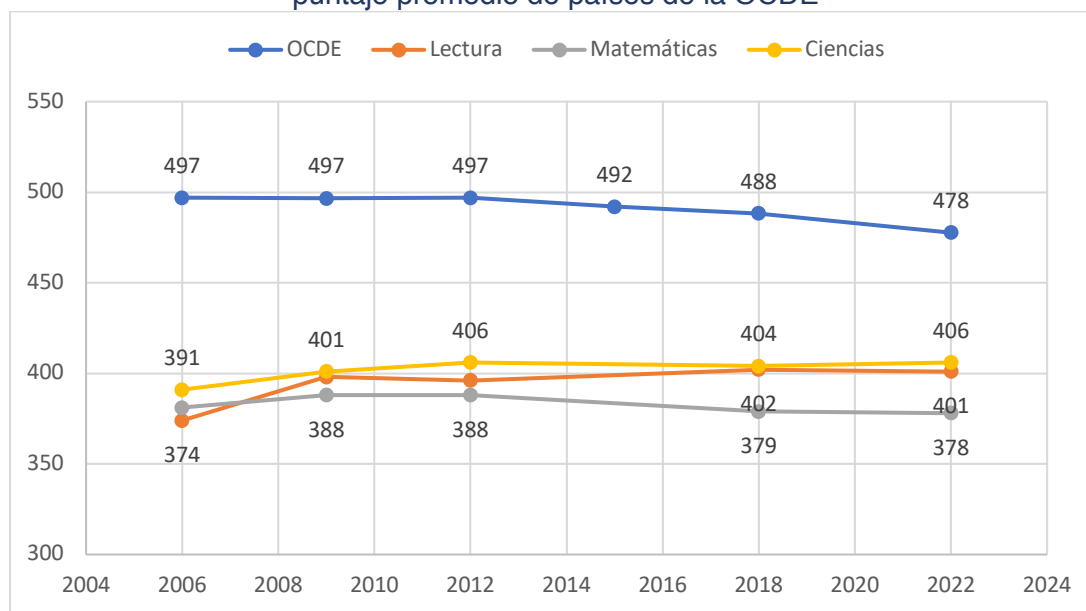
Además de la muestra nacional, Argentina incluyó tres regiones adjudicadas —Ciudad Autónoma de Buenos Aires (CABA), Córdoba y Mendoza— que participaron con resultados representativos a nivel jurisdiccional. Esto otorga un doble valor: por un lado, la comparación internacional a nivel país; y, por otro, evidencia subnacional que permite observar brechas internas. El Ministerio de Educación acompañó este proceso mediante provisión de computadoras y capacitación de tutores, lo que facilitó la logística de la aplicación.

Un aspecto metodológico relevante fue la construcción del marco muestral a partir de información estadística del Relevamiento Anual 2020, entregada a Westat en enero de 2022. Esto implicó que la selección de escuelas se basara en datos con cierto rezago

temporal respecto a la aplicación (septiembre–octubre 2022), lo que pudo generar desajustes frente a cambios en la matrícula o en la composición estudiantil. Sin embargo, el uso de listados actualizados de estudiantes por cada jurisdicción permitió mitigar parcialmente ese desfase.

En términos de resultados, Argentina mostró estabilidad relativa respecto a 2006: matemáticas (371 → 378), lectura (374 → 401) y ciencias (391 → 406). No obstante, al compararse con 2018, se evidenciaron descensos y, en el plano internacional, sus puntajes se ubicaron por debajo del promedio OCDE. El análisis por subgrupos reveló brechas importantes: en matemáticas, los hombres superaron a las mujeres en 11 puntos, mientras que los estudiantes repetidores obtuvieron 63 puntos menos que aquellos en edad teórica. Estas diferencias sugieren que factores estructurales del sistema educativo (repitencia, inequidades de género) inciden de manera más fuerte en el rendimiento que el simple diseño de la muestra.

Gráfica 2. Resultados de Argentina en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

La estrategia muestral de Argentina buscó equilibrar representatividad nacional y subnacional, lo que aporta un panorama más detallado, pero también introduce heterogeneidades internas que complejizan la interpretación de los promedios. La inclusión de jurisdicciones como CABA, Córdoba y Mendoza —con contextos socioeconómicos relativamente más favorables— probablemente atenuó las brechas que surgirían si la muestra se limitara a regiones más rezagadas. En contraste, el uso de datos de matrícula de 2020 para definir la muestra nacional pudo afectar la precisión al no captar del todo el impacto de la pandemia en la escolarización y en la composición de los grupos etarios.

Los resultados sugieren que, si bien Argentina logró mantener cierta estabilidad en el tiempo, la presencia de brechas por género y repitencia refleja desigualdades persistentes que el muestreo por sí solo no explica, pero sí visibiliza. La comparación con la OCDE muestra un desempeño bajo, aunque la relativa estabilidad frente a caídas más pronunciadas en otros países puede interpretarse como un signo de resiliencia parcial.

En el caso argentino, la forma en que fue diseñada y aplicada la muestra influyó en el tipo de resultados observados más que en su magnitud absoluta. La supervisión internacional y el esquema de reemplazo garantizan representatividad técnica, pero el rezago en los datos de base y la heterogeneidad regional introducen matices que deben tenerse en cuenta. Así, los promedios nacionales reflejan no solo el desempeño académico, sino también las tensiones entre jurisdicciones con distintos niveles de ventaja. En consecuencia, los resultados de PISA 2022 en Argentina no pueden analizarse sin considerar que la estrategia de muestreo simultáneamente asegura comparabilidad y, al mismo tiempo, suaviza desigualdades que probablemente son más profundas de lo que los promedios nacionales dejan ver.

Uruguay

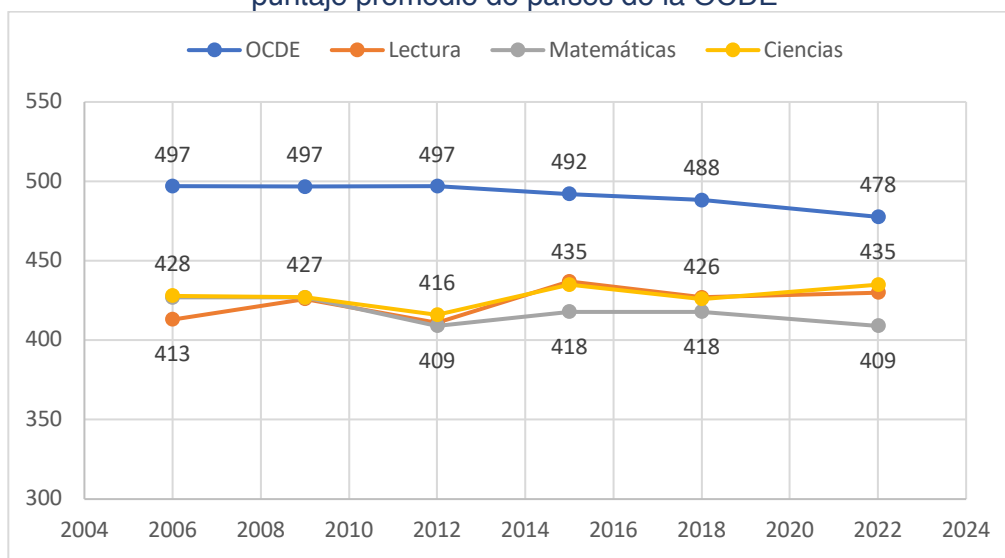
En la edición 2022 de PISA, Uruguay consolidó su posición como uno de los países con mejor desempeño educativo en América Latina, tanto por la calidad técnica de su diseño muestral como por la solidez institucional del proceso de aplicación. La muestra uruguaya estuvo conformada por 6.618 estudiantes (85 % de tasa de respuesta) pertenecientes a 222 centros educativos públicos y privados, representando a una población estimada de 43.800 jóvenes de 15 años. El diseño muestral incluyó una estratificación explícita por sector institucional (público/privado) y tipo de oferta educativa, así como una estratificación implícita por género y área geográfica. El procedimiento, desarrollado en dos fases, contempló centros de reemplazo para garantizar la representatividad en caso de inasistencias o incumplimientos.

El procedimiento de selección se desarrolló en dos fases: en la primera, se sortearon 232 centros educativos de un universo de 770, con centros de reemplazo para garantizar la representatividad; en la segunda, se seleccionaron aleatoriamente 7.798 estudiantes, de los cuales 6.618 constituyeron finalmente la muestra efectiva, tras las exclusiones por cambios de institución, necesidades educativas especiales o inasistencias reiteradas. Un aspecto distintivo fue que el 72 % de los participantes cursaba el primer año de educación media superior, lo que coincide con la edad normativa de 15 años en el país. Al analizar la distribución territorial se observa que, en el sector público, casi la mitad de los estudiantes se concentró en Montevideo, un 30% en capitales departamentales y el resto en localidades menores; mientras que en el sector privado la concentración en la capital fue aún mayor (81 %). Este patrón refleja la centralización del sistema educativo uruguayo y la fuerte dependencia de Montevideo como polo educativo.

En términos de trayectorias escolares, los datos de PISA 2018 —aún válidos como antecedente estructural— mostraban que un 64 % de los estudiantes se encontraba en el grado “normativo” (10° u 11°), pero que persistían altas tasas de repetición: 20 % había repetido en primaria y 12 % en educación media. Estas características influyen directamente en el rendimiento promedio, ya que la repitencia tiende a asociarse con menores desempeños en pruebas estandarizadas.

Los resultados de PISA 2022 evidencian una relativa estabilidad: 409 puntos en matemáticas, 430 en lectura y 435 en ciencias. En comparación con 2018, se observó una caída de 9 puntos en matemáticas, mientras que en ciencias hubo una mejora de 10 puntos, manteniendo a Uruguay entre los países con mejor desempeño de América Latina.

Gráfica 3. Resultados de Uruguay en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

El caso uruguayo muestra cómo un muestreo técnicamente robusto y balanceado entre sector público y privado permite obtener un panorama representativo del país (cobertura escolar cercana al 85% de la población de 15 años), pero al mismo tiempo evidencia la centralización educativa en Montevideo y las dificultades vinculadas a la repitencia. En términos de desempeño, Uruguay logra mantenerse como uno de los sistemas mejor posicionados en América Latina, aunque con señales de alerta en matemáticas. En este sentido, los resultados reflejan la tensión entre la fortaleza institucional del sistema y los desafíos estructurales que limitan su mejora sostenida.

República Dominicana

En el caso de República Dominicana, el diseño muestral de PISA resulta especialmente cuestionable debido a la limitada disponibilidad de información actualizada. En el informe de 2022 se hace referencia al diseño utilizado en 2015 y 2018, lo que evidencia cierta opacidad en la documentación pública más reciente. En esos años, la estrategia de

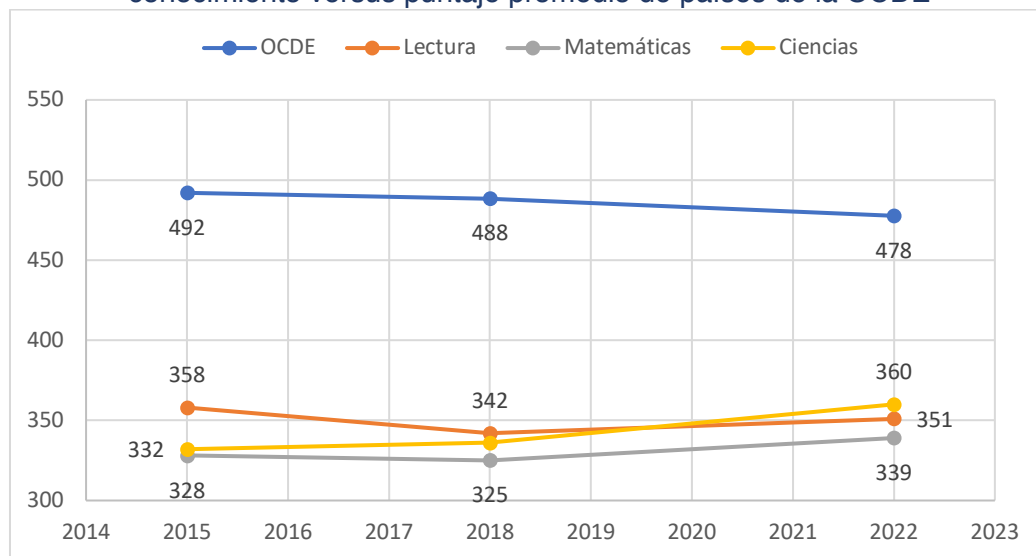
aplicación consistió en evaluar a un máximo de 35 estudiantes por centro que cumplieran con los criterios de edad y estuvieran activos en el sistema educativo.

En 2018, la muestra nacional estuvo conformada por 5.674 estudiantes de 235 centros educativos, de los cuales un 84,6 % pertenecía a liceos y escuelas públicas y un 15,4 % a colegios privados, con una distribución equilibrada por género (50,9 % mujeres y 49,1 % hombres). Esta muestra representaba apenas un 0,24 % de la población total de estudiantes, aunque aseguraba diversidad territorial al incluir centros de zonas rurales, urbanas, turísticas y marginales. A partir de ese año, la aplicación se realizó de forma digital. Para 2022, según información de la página del Ministerio de Educación, la cobertura aumentó a 6.868 estudiantes de 254 centros (208 públicos y 46 privados), representando a unos 121.900 jóvenes de 15 años a nivel nacional.

El procedimiento mantuvo la lógica bietápica probabilística: primero, la selección de centros en función de la matrícula de estudiantes en edad objetivo (15-16 años con al menos tres meses completos de edad); luego, la selección de hasta 35 estudiantes por centro que cumplieran los requisitos. Además, se incorporaron variables de caracterización como grado modal, asistencia a educación inicial, edad de inicio en primaria, repitencia escolar, nacionalidad, lengua hablada en el hogar y educación parental. Esto permitió relacionar los desempeños con factores socioeconómicos y trayectorias escolares, ofreciendo una mirada más contextualizada de los resultados.

En términos de desempeño, República Dominicana alcanzó en 2022 351 puntos en lectura, 360 en ciencias y 339 en matemáticas. Si bien se observa una mejora respecto a 2018 en todas las áreas —particularmente en matemáticas, que recuperó niveles similares a 2015—, la proporción de estudiantes bajo el nivel 2 de competencia se mantuvo prácticamente inalterada. Esto significa que, aunque el puntaje promedio muestra un avance, los rezagos más críticos de aprendizaje persisten, sin una reducción significativa en la población en situación de mayor vulnerabilidad educativa.

Gráfica 4. Resultados de República Dominicana en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

En síntesis, el diseño muestral de República Dominicana, aunque metodológicamente correcto, presenta limitaciones de cobertura que afectan su potencial comparativo. La inclusión de solo un 64 % de la población objetivo reduce la capacidad del estudio para ofrecer un diagnóstico nacional plenamente representativo (versus casos de Chile y Perú que superan el 80%). Así, más que una fotografía precisa del rendimiento individual de los estudiantes, los resultados de PISA en este país deben interpretarse como un reflejo de los desafíos estructurales de su sistema educativo: predominio del sector público, persistencia de desigualdades territoriales, y dificultades para garantizar trayectorias escolares continuas y equitativas.

Chile

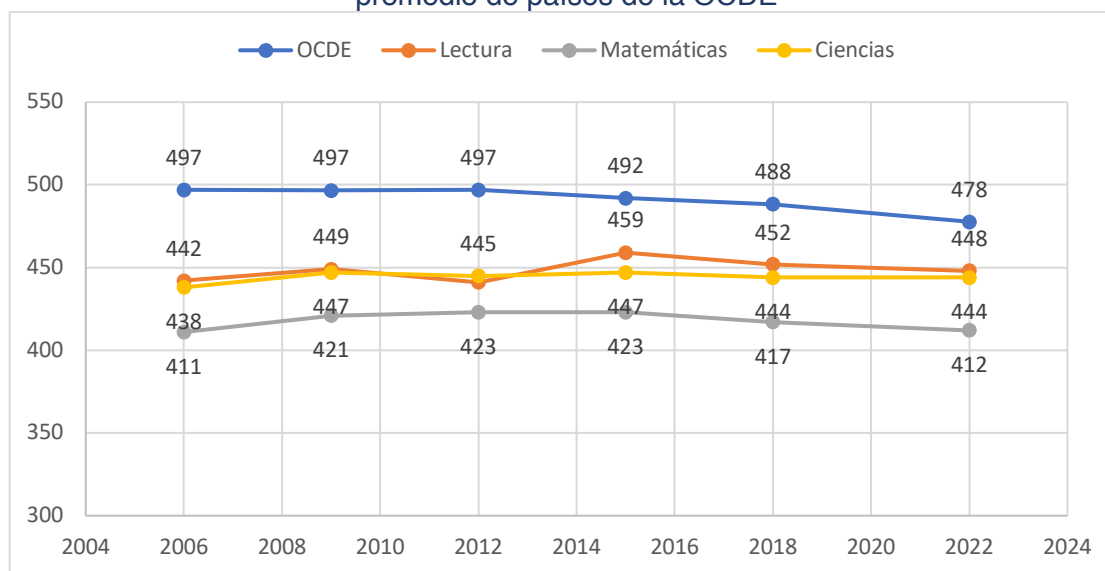
En 2022, la aplicación de PISA fue coordinada por la Agencia de Calidad de la Educación y se desarrolló entre el 13 de septiembre y el 14 de octubre en formato digital, con sesiones de dos horas. También se mantuvo una versión en papel para las preguntas de tendencia. La evaluación incluyó a estudiantes de entre 15 años y 3 meses y 16 años y 2 meses, con al menos seis años de escolaridad formal. Por primera vez, participaron estudiantes con Necesidades Educativas Especiales (NEE) integrados en escuelas regulares, quienes rindieron una versión adaptada de una hora. Sus resultados se incorporaron a la escala general de PISA, ampliando la representatividad del estudio. La incorporación por primera vez de este grupo en 2022 refuerza la equidad en la evaluación, pero también puede incidir levemente en los promedios, dado que estos estudiantes enfrentan condiciones de aprendizaje más complejas. Sin embargo, este cambio aporta una visión más completa de la realidad educativa del país.

En términos muestrales, Chile cumplió rigurosamente con las especificaciones internacionales de PISA. De los 273 establecimientos originalmente seleccionados, 209

aceptaron participar y, con los reemplazos, se alcanzó a 234 centros. De estos, 27 fueron declarados inelegibles por no contar con estudiantes de la edad objetivo, quedando 246 programados para la aplicación. Finalmente, la prueba se implementó en 230 establecimientos, lo que representa un 93 % de cobertura institucional. A nivel estudiantil, de 7.811 jóvenes seleccionados, 7.559 fueron considerados válidos y 6.488 participaron efectivamente, alcanzando una tasa de participación del 86 % de la muestra válida. Esta cobertura asegura representatividad nacional tanto a nivel regional como según modalidades y dependencias escolares.

En cuanto a los resultados, Chile obtuvo 448 puntos en lectura, 412 en matemáticas y 444 en ciencias. En comparación con 2018, los puntajes se mantuvieron relativamente estables, pero siguen por debajo de los niveles alcanzados en 2015, especialmente en lectura, donde el país había logrado uno de sus mejores desempeños históricos. En ciencias, los resultados se han mantenido prácticamente invariables desde 2006, sin una tendencia clara de mejora o deterioro. En matemáticas, la trayectoria evidencia un retroceso sostenido desde el máximo de 423 puntos alcanzado en 2012. Además, la proporción de estudiantes con puntajes por debajo del nivel mínimo de competencia (Nivel 2) no experimentó cambios significativos respecto a 2012, lo que revela dificultades persistentes para fortalecer los aprendizajes fundamentales en toda la cohorte.

Gráfica 5. Resultados de Chile en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

En síntesis, la aplicación de PISA 2022 en Chile se distingue por la solidez metodológica de su diseño muestral y por la incorporación de estudiantes con NEE, lo que le otorga mayor legitimidad y amplitud como diagnóstico nacional. Sin embargo, los resultados muestran una preocupante combinación de estabilidad y estancamiento: pese a los avances en equidad y representatividad, los aprendizajes no logran mejorar sostenidamente y, en el caso de matemáticas y lectura, muestran una tendencia

descendente desde 2015. El desafío central para el país es, por tanto, cómo traducir los esfuerzos en inclusión y cobertura en una mejora efectiva de la calidad educativa, de modo que la ampliación de la participación en PISA se acompañe de progresos sustantivos en los aprendizajes de todos los estudiantes.

Perú

En 2022, la aplicación de PISA en Perú estuvo a cargo de la Oficina de Medición de la Calidad de los Aprendizajes (UMC). La prueba se aplicó íntegramente en formato digital e incluyó a estudiantes de entre 15 años y 3 meses y 16 años y 2 meses, matriculados en la educación básica regular, en el ciclo avanzado de la Educación Básica Alternativa o en Centros de Educación Técnico-Productiva. La UMC garantizó que los centros participantes contaran con las condiciones necesarias para la aplicación, lo que refuerza la validez de los resultados obtenidos.

El país puso especial énfasis en asegurar la representatividad de la muestra frente a la población total de 15 años, estimada en 578.489 adolescentes. De este total, 536.459 (92,7 %) cumplían los criterios de la OCDE. Se excluyó un 3,36 % de la población objetivo —483 no hispanohablantes, 1.665 con discapacidad funcional o intelectual y 15.867 de zonas de difícil acceso—, lo que se mantiene por debajo del límite del 5 % fijado por la OCDE, garantizando la validez comparativa de los resultados.

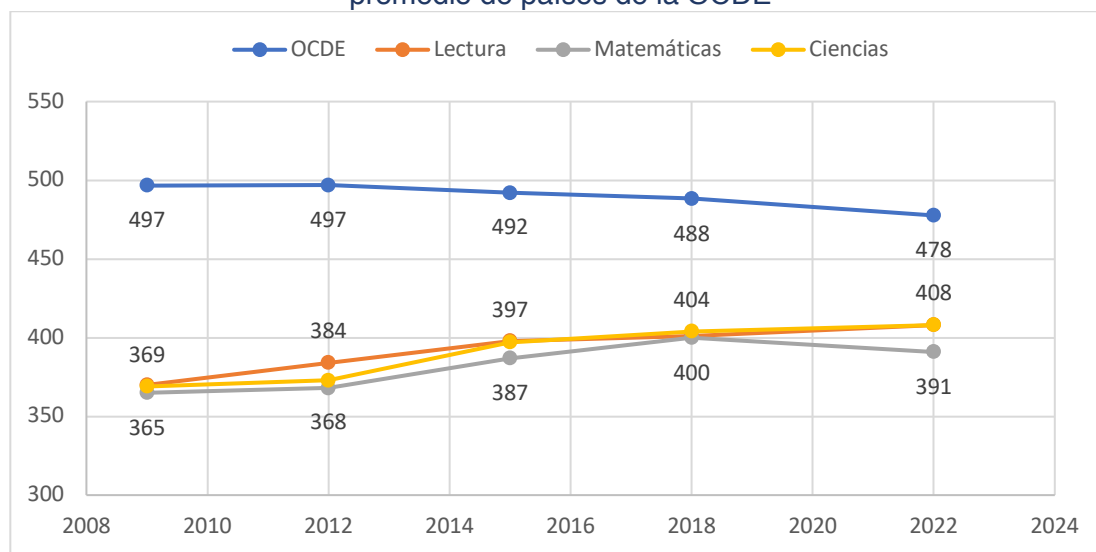
La muestra final incluyó 8.787 estudiantes de 337 escuelas, de los cuales 6.968 rindieron efectivamente las pruebas de Lectura, Matemáticas, Ciencias, Pensamiento Creativo y Educación Financiera. Esto representó el 93 % de la población escolar que cumplía los criterios y el 86,3 % de la población total de 15 años, posicionando a Perú entre los países latinoamericanos con mayor cobertura en PISA, junto con Chile. Además, esta cobertura aumentó en más de 13 puntos porcentuales respecto a 2018, lo que evidencia una mejora significativa en la capacidad del sistema para incluir a un mayor número de jóvenes en la evaluación.

En cuanto a la trayectoria escolar, la mayoría de los estudiantes evaluados cursaba 4° de secundaria (70,7 %), mientras que el 12,7 % se encontraba en 3° y un 13,4 % en 5°. Esta heterogeneidad refleja fenómenos de sobreedad y rezago escolar, lo cual influye directamente en los resultados, ya que los adolescentes de la misma edad enfrentan distintos niveles de exposición curricular. Un dato relevante es la disminución en la tasa de repetición: un 13,5 % reportó haber repetido al menos un grado, lo que supone una reducción en comparación con ciclos anteriores y puede asociarse a las políticas de promoción automática aplicadas durante la pandemia.

En cuanto al desempeño, Perú obtuvo 408 puntos en lectura, 391 en matemáticas y 403 en ciencias. Frente a 2018, los resultados en lectura y ciencias se mantuvieron estables, mientras que matemáticas registró una leve caída (de 400 a 391). Sin embargo, la perspectiva de largo plazo (2012–2022) muestra una mejora sostenida en las tres áreas: +23 puntos en lectura, +23 en matemáticas y +19 en ciencias. Asimismo, entre 2012 y

2022, la proporción de estudiantes por debajo del nivel mínimo de competencia se redujo en 8 puntos porcentuales en matemáticas, 9 en lectura y 16 en ciencias, lo que refleja avances en la reducción de los aprendizajes más rezagados.

Gráfica 6. Resultados de Perú en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

En general, podría decirse que, según el diseño muestral, con más del 86 % de la población total de 15 años incluida, Perú logra que los resultados reflejen de manera muy cercana la realidad educativa del país, reduciendo los sesgos asociados a exclusiones por zonas o grupos poblacionales. Por otro lado, la heterogeneidad en la ubicación escolar de los estudiantes de 15 años (entre 3° y 5° de secundaria) puede incidir en los resultados, ya que el nivel de exposición curricular varía según el grado. Este aspecto probablemente explica parte de la dispersión en los puntajes y resalta la importancia de atender la sobreedad y los rezagos escolares.

La consistencia de la muestra a lo largo de las aplicaciones permite afirmar que las mejoras observadas no responden a cambios metodológicos, sino a progresos reales en aprendizajes básicos. En síntesis, Perú se consolida como un caso de mejora sostenida y de amplia representatividad regional en PISA, combinando un diseño muestral robusto con un progreso educativo que, aunque aún distante de los promedios de la OCDE, evidencia un impacto positivo en el desempeño académico.

México

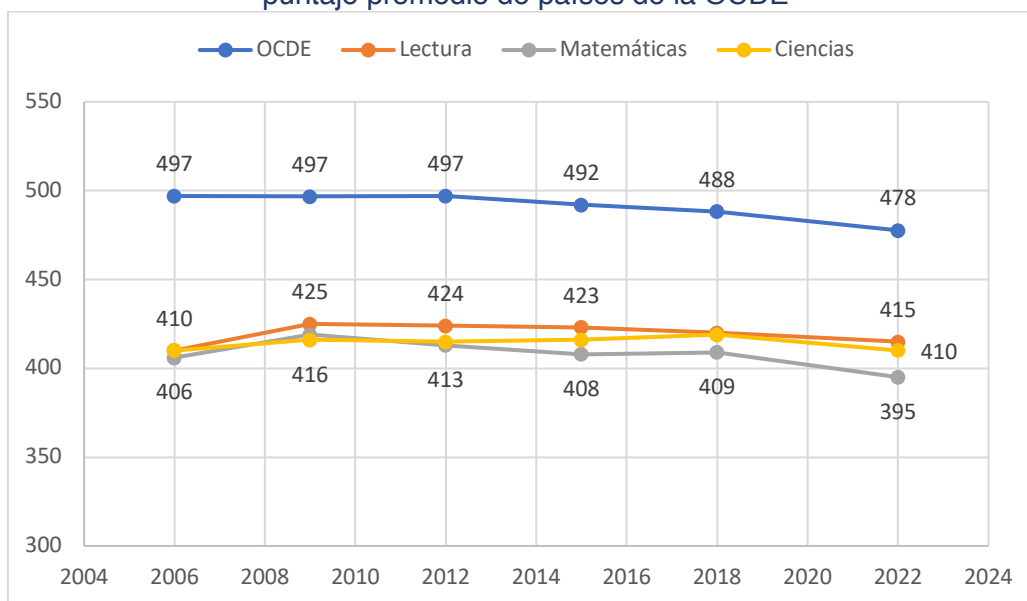
La información operativa más detallada disponible proviene de 2015, cuando el Instituto Nacional para la Evaluación de la Educación (INEE) organizó la aplicación de PISA. Ese año fue la primera edición implementada en formato digital, lo que significó una transición importante pero también limitó la capacidad de análisis regional, dado que el diseño se centró en la representatividad nacional.

En 2015 participaron 7.568 estudiantes de 275 escuelas, representando a 1.740.544 adolescentes de 15 años. El 86 % estaba matriculado en escuelas públicas y el 88 % en zonas urbanas. En términos de trayectorias escolares, el 69,3 % cursaba educación media superior y el 30,7 % secundaria, con predominio de modalidades de bachillerato general y tecnológico. La heterogeneidad en los niveles de estudio refleja la compleja estructura del sistema educativo mexicano, donde la progresión escolar de los estudiantes de 15 años varía según región, modalidad y condición socioeconómica. Cabe destacar que estados como Michoacán y Oaxaca tuvieron tasas de participación escolar menores, lo que advierte desigualdades territoriales en cobertura.

En 2022, participaron 6.288 estudiantes de 280 escuelas, representando a unos 1.393.700 estudiantes de 15 años. La muestra cubrió únicamente al 64 % de la población total de esa edad, lo que coloca a México entre los países latinoamericanos con menor cobertura en PISA. Esta reducción implica un mayor riesgo de subrepresentación de grupos sociales o territoriales, especialmente aquellos en contextos de marginación o de difícil acceso.

En 2022, México obtuvo 415 puntos en lectura, 395 en matemáticas y 410 en ciencias. En comparación con 2018, lectura se mantuvo estable, mientras que matemáticas y ciencias disminuyeron significativamente, regresando a niveles similares a los observados entre 2003 y 2006. La evolución de largo plazo muestra una tendencia preocupante: desde 2012, el porcentaje de estudiantes por debajo del nivel mínimo aumentó en 11 puntos en matemáticas y en 5 puntos en lectura, mientras que ciencias permaneció prácticamente estable.

Gráfica 7. Resultados de México en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

En general, puede decirse que la menor representatividad de la población de 15 años respecto a países vecinos (Perú 86,3 %, Chile 86,5 %) implica que los resultados pueden no reflejar plenamente la diversidad del sistema educativo mexicano, en especial las zonas rurales, marginadas o con menor participación escolar. Por otro lado, se observa que la muestra en 2015 estuvo fuertemente concentrada en escuelas urbanas, lo que puede reducir la visibilidad de brechas urbano-rurales, indígenas o regionales, que en México son especialmente significativas. Finalmente, la baja participación de ciertos estados en 2015, sumada a la menor cobertura en 2022, sugiere que las regiones con mayores rezagos educativos pueden estar subrepresentadas, lo que atenúa la magnitud de las brechas en los resultados nacionales.

En conclusión, México enfrenta un doble desafío: por un lado, la caída sostenida en matemáticas y ciencias que lo ha llevado de vuelta a los niveles de inicios de los 2000; y por otro, un diseño muestral con cobertura limitada que restringe la lectura precisa de desigualdades internas.

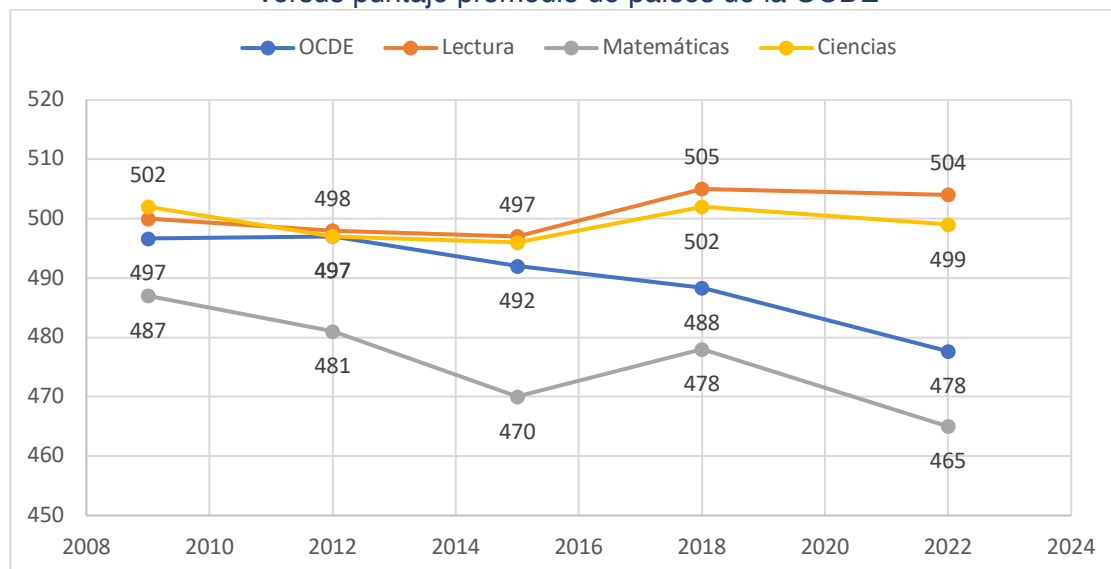
Estados Unidos

En 2022, la muestra de Estados Unidos fue diseñada por el National Center for Education Statistics (NCES) siguiendo criterios de estratificación explícita por región (Noreste, Medio Oeste, Sur y Oeste) y tipo de escuela (pública o privada), complementados con cinco variables implícitas: grado escolar, ubicación geográfica (urbana, suburbana, rural), composición étnica, género y estado. La muestra nacional incluyó 297 escuelas —casi el doble del requisito internacional de 150— con el fin de reducir el sesgo por falta de respuesta. En cada escuela se seleccionaron 52 estudiantes de 15 años, de los cuales 41 realizaron las pruebas principales (matemáticas, lectura y ciencias) y 11 la prueba opcional de competencia financiera. En total participaron 4.552 estudiantes en la evaluación principal y 1.109 en alfabetización financiera.

Un aspecto distintivo de la muestra estadounidense fue la implementación de adaptaciones para estudiantes con necesidades educativas especiales mediante el instrumento Une Heure (UH), que incluía menos ítems y mayor flexibilidad de tiempo. Esta medida buscó reducir las tasas de exclusión y asegurar la comparabilidad internacional, reflejando a la vez un compromiso explícito con la equidad y la inclusión.

Los resultados de PISA 2022 mostraron un puntaje de 465 en matemáticas, 504 en lectura y 499 en ciencias. En el largo plazo, Estados Unidos presenta una tendencia descendente en matemáticas (–22 puntos entre 2009 y 2022), mientras que en lectura y ciencias ha mantenido relativa estabilidad, con una ligera mejora en lectura respecto a 2015. Pese a la caída en matemáticas, el país mejoró su posición relativa frente a otros sistemas educativos debido a que los descensos globales tras la pandemia fueron más pronunciados en naciones comparables.

Gráfica 8. Resultados de Estados Unidos en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

La estrategia muestral estadounidense refleja un esfuerzo deliberado por asegurar representatividad en un sistema educativo altamente heterogéneo, tanto por la diversidad étnica y socioeconómica como por la organización federal. El hecho de que la muestra incluya un número significativamente mayor de escuelas que el mínimo internacional reduce la probabilidad de sesgos de selección, aunque al mismo tiempo plantea un reto para capturar de manera precisa las desigualdades internas. La inclusión de estudiantes con necesidades especiales mediante instrumentos diferenciados representa un avance en equidad, pero también introduce complejidades metodológicas que pueden influir en la interpretación de los resultados agregados.

En cuanto a los desempeños, la persistente brecha en matemáticas podría estar vinculada a factores estructurales del sistema, como la segmentación por estados y distritos, la diversidad curricular y las desigualdades socioeconómicas, más que al diseño muestral en sí mismo. No obstante, el muestreo estratificado por variables étnicas, de ubicación y género asegura que los resultados reflejen con mayor fidelidad la realidad de un país profundamente diverso, evitando una sobre-representación de contextos urbanos o de escuelas privadas, como ocurre en otras naciones. Así, la experiencia estadounidense evidencia que un diseño muestral sólido no necesariamente se traduce en mejores resultados, pero sí en diagnósticos más confiables para orientar políticas educativas.

Inglaterra

En 2022, Inglaterra seleccionó inicialmente 201 escuelas para PISA, considerando variables como el tipo de establecimiento (mixto o de un solo género), la región geográfica y el rendimiento académico previo. También se aplicó un criterio de probabilidad

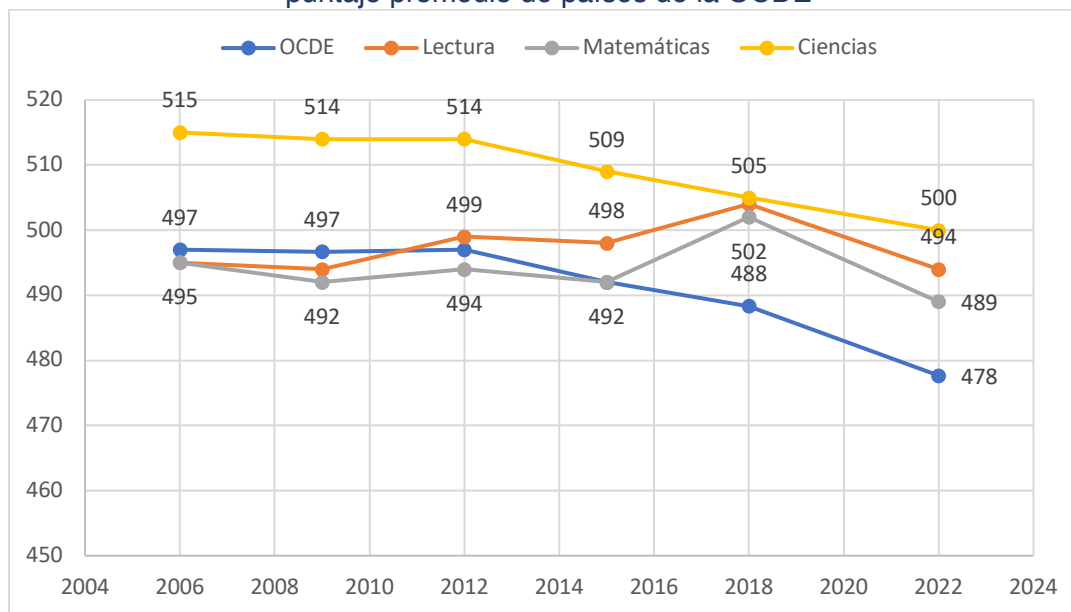
proporcional al tamaño, lo que otorgó mayor peso a las escuelas con más estudiantes de 15 años. Con el fin de mitigar posibles problemas de participación, se identificaron dos centros de reemplazo por cada escuela de la muestra principal. Sin embargo, los desafíos logísticos y de calendario (noviembre–diciembre de 2022) redujeron la participación: la muestra final incluyó 165 escuelas, con una tasa ponderada del 66 % en la primera fase y del 82 % tras incorporar reemplazos, cifra inferior al objetivo internacional (94,3 %). A nivel estudiantil, la tasa de respuesta fue del 75 %, igualmente por debajo del umbral mínimo de la OCDE (80 %). En total, participaron 4.763 alumnos de 15 años.

El incumplimiento de los estándares técnicos de PISA obligó a realizar un Análisis de Sesgo por Falta de Respuesta (ANFR). Los hallazgos mostraron una sobrerrepresentación de estudiantes con alto rendimiento previo y una subrepresentación de alumnos elegibles para comidas escolares gratuitas, indicador clave de desventaja socioeconómica. Como resultado, tanto los puntajes como el índice socioeconómico (ESCS) podrían estar sesgados al alza, en un rango estimado por la OCDE de 7 a 8 puntos.

Pese a ello, la muestra mantuvo diversidad institucional —tipo de escuela, composición lingüística, necesidades educativas especiales— lo que reduce riesgos de sesgos estructurales. No obstante, la sobrerrepresentación de estudiantes de alto rendimiento obliga a interpretar los resultados con cautela, en especial en comparaciones internacionales. Situaciones similares se observaron en otros países de la OCDE (EE. UU., Canadá, Países Bajos, Irlanda, entre otros), lo que refleja las dificultades de asegurar plena representatividad en el contexto de la pandemia y la fatiga evaluativa.

En cuanto al desempeño, Inglaterra obtuvo 489 puntos en matemáticas, 494 en lectura y 500 en ciencias. Estos resultados reflejan un retroceso respecto a 2018 en matemáticas y lectura, con estabilidad en ciencias. No obstante, el 76 % de los estudiantes alcanzó al menos el nivel 2 en matemáticas, superando el promedio de la OCDE y confirmando la solidez relativa del sistema frente a los descensos globales observados tras la pandemia.

Gráfica 9. Resultados de Inglaterra en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

El caso inglés ilustra cómo las limitaciones en la tasa de respuesta pueden comprometer la representatividad de la muestra, aun cuando el diseño inicial fuera riguroso. La sobrerrepresentación de estudiantes de mayor rendimiento, junto con la menor participación de estudiantes de contextos desfavorecidos, sugiere que los resultados nacionales podrían reflejar un desempeño algo más favorable del que se observaría con una muestra completamente equilibrada. En este sentido, Inglaterra evidencia la importancia de acompañar las evaluaciones internacionales con estudios de sesgo que permitan los hallazgos y fortalecer la credibilidad del diagnóstico. Asimismo, su experiencia refleja un reto común a varios países de la OCDE (como Estados Unidos, Canadá, Países Bajos e Irlanda): sostener la representatividad plena en un contexto de fatiga evaluativa y complejidades pospandemia.

España

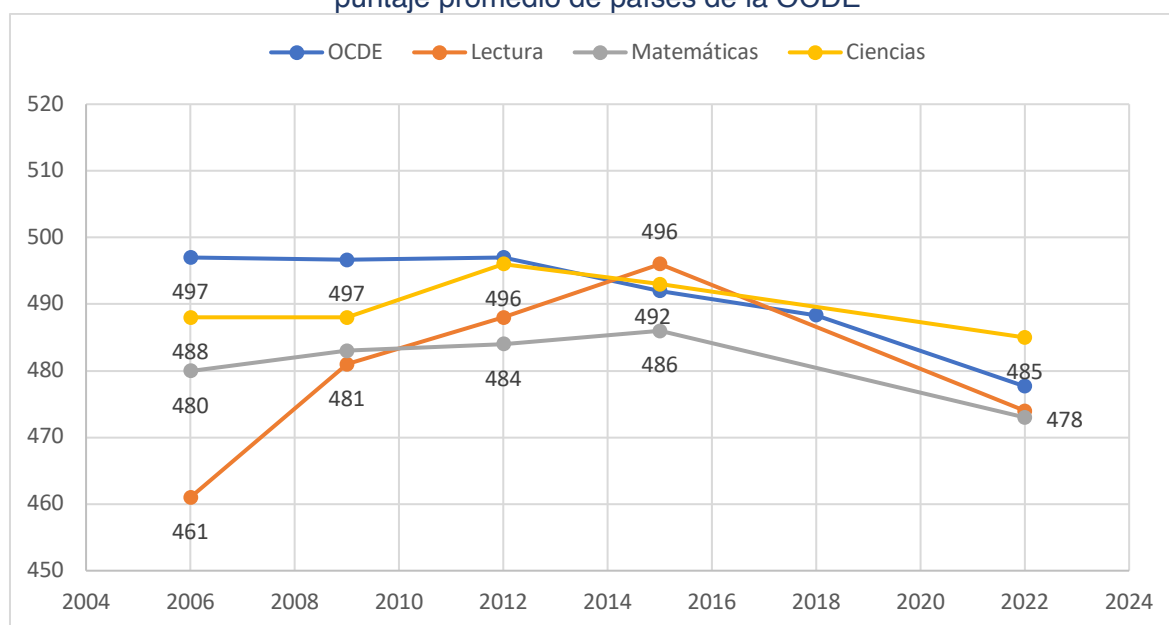
En España, los estudiantes elegibles para PISA 2022 debían tener entre 15 años y 3 meses y 16 años y 2 meses al momento de la aplicación, además de haber permanecido al menos seis años en el sistema educativo. En dicha edición participaron aproximadamente 30.800 estudiantes de 966 centros escolares, lo que representa una cobertura cercana al 90 % de la población total de 15 años en el país (aprox. 459.000 jóvenes). La muestra fue representativa en todas las comunidades y ciudades autónomas, con predominio de estudiantes matriculados en 4.º de ESO.

El diseño muestral aplicado siguió un esquema bietápico: en la primera etapa, se seleccionaron los centros en función de su tamaño y de características previamente estratificadas para maximizar la precisión de los estimadores; en la segunda, se escogió

aleatoriamente a los estudiantes dentro de los centros seleccionados, con inclusión completa en aquellos de menor tamaño. Sin embargo, a diferencia de otros países, no se dispone de información pública detallada y actualizada sobre los criterios técnicos específicos del muestreo, lo que limita la posibilidad de un escrutinio independiente sobre el grado de representatividad alcanzado.

En cuanto a los resultados, España registró en 2022 un descenso en las tres áreas evaluadas respecto a 2015, con especial retroceso en matemáticas, donde alcanzó 474 puntos, el nivel más bajo desde 2006 y por debajo del promedio de la OCDE. En lectura (477) y ciencias (485), los resultados fueron semejantes a los de 2018, aunque inferiores a los de 2015. La evolución de largo plazo muestra un aumento sostenido de estudiantes por debajo del nivel 2 de competencia en matemáticas, lectura y ciencias —considerado el umbral mínimo para desenvolverse en la vida adulta—, acompañado de una reducción de los alumnos en los niveles de excelencia (5 y 6), especialmente en matemáticas.

Gráfica 10. Resultados de España en las pruebas PISA según área de conocimiento versus puntaje promedio de países de la OCDE



Fuente: OECD (2023), PISA 2022 Results – Country Notes.

Este patrón refleja un ensanchamiento de la brecha de aprendizajes y un debilitamiento en la formación de competencias avanzadas, lo que supone un riesgo tanto para la equidad como para la competitividad del sistema educativo español. Aunque la muestra garantiza representatividad geográfica y por tipo de centro, la falta de mayor transparencia en los detalles técnicos del muestreo limita la posibilidad de identificar sesgos asociados a factores socioeconómicos o institucionales.

En síntesis, el caso español muestra una paradoja: un diseño muestral amplio y representativo en términos territoriales, pero acompañado de resultados que evidencian

una pérdida de aprendizajes básicos y avanzados en la última década. El reto consiste en revertir el deterioro sostenido en matemáticas y frenar la expansión de la proporción de estudiantes con bajo rendimiento, al tiempo que se recupera la capacidad del sistema para promover la excelencia académica. La experiencia sugiere que el desafío no radica únicamente en la cobertura o representatividad de la muestra, sino en la capacidad del sistema para sostener mejoras a largo plazo frente a factores estructurales y coyunturales, incluida la pandemia.

6. Propuestas para mejorar el diseño muestral de PISA

El análisis de los resultados y limitaciones identificadas en la aplicación de PISA sugiere la necesidad de continuar los esfuerzos para avanzar hacia un diseño muestral más inclusivo y comparable entre países. Algunas estrategias orientadas a mejorar la validez y representatividad de la evaluación:

- Publicación de los procesos de selección muestral

Para todas las pruebas estandarizadas que pasan por un proceso muestral, es una buena práctica publicar los criterios, fases y decisiones metodológicas adoptadas durante la selección de escuelas y estudiantes. Esto incluye la difusión de los parámetros de estratificación, los motivos de exclusión de centros o estudiantes y los ajustes realizados en campo. La apertura de esta información fortalecería la confianza de la comunidad académica y permitiría encontrar opciones de mejora por parte de personas locales o externos.

- Homologación internacional de las tasas de exclusión con mecanismos de control

Las tasas de exclusión deben estar sujetas a un sistema homogéneo de cálculo y verificación. Se sugiere establecer umbrales máximos vinculantes, acompañados de mecanismos de ajuste estadístico en casos donde se superen dichos límites. Esto podría reducir las diferencias en la aplicación de criterios de exclusión y reduciría los sesgos en la comparabilidad de los resultados.

- Inclusión efectiva de estudiantes con necesidades educativas especiales

La representatividad de PISA podría incorporar de manera más sistemática a estudiantes con discapacidad o necesidades educativas especiales con adaptaciones metodológicas (instrumentos accesibles, apoyos tecnológicos, formatos alternativos de prueba) que permitan capturar el desempeño de estos grupos.

- Revisión de las fechas de corte escolar para reducir sesgos de cohorte

Los criterios que definen la edad y el momento de aplicación son cruciales para mitigar los sesgos derivados de diferencias en calendarios académicos y fechas de ingreso escolar de los diferentes países. Una opción es revisar si los estudiantes evaluados tienen trayectorias educativas comparables en términos de exposición curricular, y de ser muy diferentes, revisar ajustes o señalar la anotación en las bases para que los investigadores las tengan en cuenta.

- Ampliación y ajuste proporcional de los tamaños muestrales para mayor representación regional

Se recomienda que las muestras se dimensionen en función de la población estudiantil total, el peso relativo de regiones o grupos sociales. Esto permitiría capturar con mayor precisión brechas internas, especialmente en contextos con alta desigualdad territorial.

- Complementariedad con encuestas de hogares para integrar a poblaciones no escolarizadas

Para abordar la invisibilidad de adolescentes fuera del sistema educativo, sería ideal implementar un módulo paralelo de encuestas de hogares en una submuestra representativa. Esta estrategia ampliaría el alcance de PISA hacia quienes no asisten a la escuela, generando un panorama más integral de los aprendizajes, las barreras de acceso y las desigualdades estructurales. También permitiría proyectar los posibles sesgos de medición asociados a quienes estén fuera del sistema.

- Fortalecimiento de los mecanismos de monitoreo independiente

Además de las auditorías técnicas realizadas por la OCDE, sería pertinente habilitar espacios de revisión independiente por parte de organismos académicos y centros de investigación. Esto contribuiría a la identificación temprana de sesgos y a la construcción de consensos internacionales sobre posibles prácticas adicionales que se pudieran considerar en cuanto al diseño muestral.

Referencias bibliográficas

Anders, J., Has, S., Jerrim, J., Shure, N., & Zieger, L. (2020). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Educational Assessment, Evaluation and Accountability*, 33(1), 229–249. <https://doi.org/10.1007/s11092-020-09329-5>

Andersson, C., & Sandgren Massih, S. (2023). PISA 2018: Did Sweden exclude students according to the rules? *Assessment in Education: Principles, Policy & Practice*, 30(1), 33–52. <https://doi.org/10.1080/0969594X.2023.2189566>

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., & Blossfeld, H. P. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions (Vol. 14). VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/s11618-011-0181-8>

Bank, V. (2012). On OECD policies and the pitfalls in economy-driven education: The case of Germany. *Journal of Curriculum Studies*, 44(2), 193-210. doi: <http://dx.doi.org/10.1080/00220272.2011.639903>

Cai, T. (2013). Investigation of ways to handle sampling weights for multilevel model analyses. *Sociological Methodology*, 43(1), 178–219. <https://doi.org/10.1177/0081175012460221>.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). A Wiley publication in applied statistics.

Daniel Caro & Leonidas Kyriakides (2019) Assessment design and quality of inferences in PISA: limitations and recommendations for improvement, *Assessment in Education: Principles, Policy & Practice*, 26:4, 363-368, DOI: 10.1080/0969594X.2019.1645990

De Witte, K. & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17), 2401-2412. doi: <http://dx.doi.org/10.1080/00036846.2012.665602>

Durrant, G.B. and Schnepf, S.V. (2018), Which schools and pupils respond to educational achievement surveys?: a focus on the English Programme for International Student Assessment sample. *J. R. Stat. Soc. A*, 181: 1057-1075. <https://doi.org/10.1111/rssa.12337>

Fernandez-Cano, Antonio (2016). Una crítica metodológica de las evaluaciones PISA. *RELIEVE*, 22(1), art. M15. DOI: <http://dx.doi.org/10.7203/relieve.22.1.880>

Fuchs, Thomas; Woessmann, Ludger (2004) : What accounts for international differences in student performance?, CESifo Working Paper, No. 1235, Center for Economic Studies and ifo Institute (CESifo), Munich

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK?, Review of Education, 9(3). <https://doi.org/10.1002/rev3.3270>

Judkins, D. R. (1990). Fay's method for variance estimation. Journal of Official Statistics, 6, 223–239.

Krohne, U., Meier, G., & Tillmann, K.-J. (2004). PISA 2000: Eine Kritik der Kritik – Anmerkungen zur Kritik an den Methoden und Ergebnissen der PISA-Studie. Zeitschrift für Erziehungswissenschaft, 7(2), 149–166. <https://doi.org/10.1007/s11618-004-0020-2>

Leonhardt, D. (2000, July 28). John Tukey, 85, Statistician; Coined the Word 'Software'. The New York Times. <https://www.nytimes.com/2000/07/28/us/john-tukey-85-statistician-coined-the-word-software.html>

LeRoy, B.-W., Samuel, P., Deluca, M., & Evans, P. (2018). Students with special educational needs within PISA. Assessment in Education: Principles, Policy & Practice. doi:10.1080/0969594X.2017.1421523

Mang, J. (2023). Evaluating the PISA sampling design by stratification schemes and weighting procedures in hierarchical modelling [Tesis doctoral, Ludwig-Maximilians-Universität München]. Universitätsbibliothek der LMU München. <https://doi.org/10.5282/edoc.32772>

Meinck, S. (2020). Sampling, Weighting, and Variance Estimation. In H. Wagemaker (Ed.), IEA Research for Education, A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA). Reliability and Validity of International Large-Scale Assessment: Understanding IEA's Comparative Studies of Student Achievement (1st ed., pp. 113–129). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_7

Meinck, S., & Vandenplas, C. (2021). Sampling design in ILSA. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), International handbook of comparative large-scale studies in education: Perspectives, methods and findings (pp. 1–25). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_25-1

Micklewright, J., Schnepf, S. V., & Skinner, C. (2012). Non-response biases in surveys of schoolchildren: The case of the English Programme for International Student Assessment

(PISA) samples. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 175(4), 915–938. <https://www.jstor.org/stable/23355308>

Mortimore, P. (2009). Alternative models for analysing and representing countries' performance in PISA. Education International Research Institute.

OECD. (2017). PISA 2015 technical report. Paris: OECD Publishing.

OECD. (2023). PISA 2022 results (Volume I and II) – Country Notes: Colombia. OECD. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/colombia_dd5f34d9-en.html

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>

Reiss, K., Weis, M., Klieme, E., & Köller, O. (2019). PISA 2018. Grundbildung im internationalen Vergleich. Münster: Waxmann. <https://doi.org/10.31244/9783830991007>

Rust, K. F., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3), 283–310. <https://doi.org/10.1177/096228029600500305>

Skinner, C. J. (2014). Probability proportional to size (PPS) sampling (pp. 1–5). Wiley StatsRef: Statistical Reference Online

Spaull, N. (2018). Who makes it into PISA? Understanding the impact of PISA sample eligibility using Turkey as a case study (PISA 2003–PISA 2012). *Assessment in Education: Principles, Policy & Practice*, 26(4), 369–386. <https://doi.org/10.1080/0969594X.2018.1504742>

Valliant, R., Dever, J. A., & Kreuter, F. (Eds.). (2018a). *Practical Tools for Designing and Weighting Survey Samples*. Springer, Cham



El Laboratorio de Economía de la Educación (LEE) es una iniciativa de la Pontificia Universidad Javeriana que investiga, evalúa, analiza y provee información cuantitativa sobre el sistema educativo.

LEE pretende guiar la toma de decisiones, así como también el desarrollo de innovaciones y políticas educativas efectivas para impulsar la transformación de la educación en Colombia.

Si necesita citar este documento, hágalo de la siguiente manera:

Laboratorio de Economía de la Educación (LEE) de la Pontificia Universidad Javeriana. (2025). Informe No. 128. Diversidad de muestras, diversidad de resultados: lecciones internacionales desde PISA. Disponible en <https://lee.javeriana.edu.co/publicaciones-y-documentos>