



¿CÓMO INTERPRETAR LOS RESULTADOS DE PISA?



Textos de
divulgación

INEE
Instituto Nacional para la
Evaluación de la Educación
México



¿CÓMO INTERPRETAR LOS RESULTADOS DE PISA?

Este material tiene como objetivo exponer algunos elementos metodológicos del Programa para la Evaluación Internacional de los Estudiantes (PISA, por sus siglas en inglés) y, con ello, propiciar la reflexión sobre las distintas maneras de presentar los resultados del desempeño de los estudiantes de 15 años,¹ todo lo cual permitirá a los interesados contar con mayores elementos para hacer una interpretación al respecto. Se debe considerar que la pertinencia de ésta depende, en gran medida, del conocimiento que se tenga acerca del diseño y del alcance de la evaluación PISA. Adicionalmente, el uso adecuado de dicha interpretación implica entender claramente el tipo de datos que arroja la prueba, y que suelen consistir en las medias y los porcentajes de estudiantes en los niveles de desempeño.

PREGUNTAS, PUNTAJES Y DISEÑO DE LA PRUEBA

En general, cuando se reportan los resultados, se informa sobre la calidad del Sistema Educativo Nacional respecto de la preparación que han tenido los estudiantes tanto a lo largo de su formación escolarizada como también fuera del aula. Para dar a conocer ese diagnóstico del sistema educativo, PISA desarrolla instrumentos que se aplican a estudiantes de 15 años. En particular, las pruebas comprenden unidades de reactivos o preguntas sobre tres áreas o competencias: científica, lectora y matemática. Algunas de éstas requieren seleccionar la alternativa correcta entre varias opciones (mediante reactivos de opción múltiple), mientras que otras demandan, en cambio, que cada estudiante desarrolle su propia respuesta (mediante reactivos de respuesta abierta).

Las preguntas poseen un distinto grado de dificultad; por ejemplo, las de respuesta abierta no se califican sólo como correctas o incorrectas, sino que pueden ser parcialmente correctas. Por características como éstas, el puntaje total de un país no se consigue simplemente con la suma de las respuestas acertadas de sus estudiantes. Se obtiene mediante técnicas estadísticas complejas, de probada solidez, que se

¹ De manera económica se alude a los estudiantes de 15 años, pero, en sentido estricto, la edad de los participantes se ubica entre los 15 años tres meses y los 16 años dos meses.

Nota: parte del contenido de este material se incluyó originalmente en el capítulo 2 del informe de resultados México en PISA 2009, publicado en 2010.

derivan de la Teoría de Respuesta al Ítem. Con base en éstas, los resultados de PISA se presentan en escalas (una por área o competencia) que se caracterizan por tener una media teórica de 500 puntos y una desviación estándar de 100.²

Por otra parte, dado el gran número de preguntas que integran las pruebas usadas en PISA, no sería adecuado solicitar a cada estudiante que las respondiera en su totalidad. Por esta razón, la prueba se divide en partes que se aplican a distintos estudiantes. Esta distribución se realiza mediante un cuidadoso diseño denominado *matricial* que no permite dar resultados de cada estudiante o de cada escuela, pero sí del conjunto de estudiantes de un país o región.

En cualquier evaluación hay márgenes de error que son inevitables, pero en este caso se encuentran dentro de un espectro aceptable. Dichos márgenes siempre son dados a conocer junto con los resultados.

LOS ORDENAMIENTOS SIMPLES O RANKINGS

Una forma de presentar los resultados de PISA se reduce a hacer un ordenamiento simple de los países con base en la media de las puntuaciones obtenidas. Estos ordenamientos, conocidos como *rankings*, suelen resultar atractivos porque permiten obtener conclusiones sobre cuáles sistemas educativos son de mejor o peor calidad en función del lugar que ocupen. Estas interpretaciones, sin embargo, carecen de sustento sólido, pues no toman en cuenta varios aspectos como los que se resumen a continuación:

- **Multidimensionalidad de las pruebas.** Los resultados de cada país en las tres competencias (lectora, matemática y científica) suelen situarse en un rango cercano; dado que la coincidencia no es perfecta, las puntuaciones de cada escala dan lugar a ordenamientos diferentes. Además, una misma competencia puede tener varias dimensiones y cada una puede dar lugar a una subescala y a un ordenamiento distinto.
- **Distancia entre puntuaciones.** Los *rankings* conducen a interpretaciones que ignoran la importancia de la distancia que separa cualquier pareja de puntuaciones y que puede ser pequeña o grande. Las diferencias pequeñas son engañosas, sobre todo cuando son menores al *intervalo de confianza* de la medición.

² La media de 500 y la desviación estándar de 100 se establecieron en PISA 2000 como referentes para los ciclos posteriores, en los cuales la media puede ser distinta.

Como se mencionó, los resultados de las pruebas PISA no son el promedio de las puntuaciones de todos los sujetos de la población, sino estimaciones a partir de lo obtenido por los integrantes de una muestra, por lo que hay un margen de error inevitable. Un buen diseño muestral permite estimar con precisión ese margen, con una probabilidad determinada, como se hace en PISA.

Cuando la distancia que separa los promedios obtenidos por los estudiantes de dos países es menor al margen de error definido, con cierto grado de probabilidad, no se puede afirmar que el resultado de quienes tienen la puntuación superior sea efectivamente mejor que el de los otros. En esos casos, se considera un empate técnico entre los países en cuestión.

Tomando en cuenta lo anterior; y con el propósito de evitar las interpretaciones simplistas, generalmente se presentan los resultados con ordenamientos de países de forma tal que la distancia que los separa en las gráficas es proporcional a la diferencia de las puntuaciones promedio. Además, siempre se incluye el intervalo de confianza correspondiente para que se aprecie con claridad cuáles puntuaciones no difieren de manera significativa —en términos estadísticos— porque los intervalos de confianza se traslapan, y en cuáles casos la diferencia sí es importante.



¿PUNTUACIÓN ALTA ES SINÓNIMO DE MEJOR CALIDAD EDUCATIVA?

Una limitación aún más elemental de estas interpretaciones basadas en ordenamientos simples de puntuaciones promedio es la que consiste en interpretar una puntuación más alta como reflejo inequívoco de mejor calidad de las escuelas, y una puntuación más baja como signo indudable de menor calidad. En realidad, los resultados se deben a muchos factores, unos del ámbito escolar; y otros —cuya importancia se reitera estudio tras estudio— del entorno familiar y social de los alumnos: en una escuela que funciona bien pueden obtenerse resultados inferiores a los de otra cuyo funcionamiento es menos ordenado, precisamente debido al peso de los factores del entorno.

En sociedades con alta desigualdad económica, como la mexicana, las interpretaciones simplistas de los *rankings* pueden ser injustas con las escuelas que atienden a alumnos de medios desfavorecidos, máxime si se tiene en cuenta que además suelen contar con menos recursos que las escuelas a las que asisten mayoritariamente alumnos de medios más favorables. Esto no debe llevar a la conclusión de que los resultados de pruebas como las de PISA no son relevantes, pero sí debe

alertar sobre la necesidad de tener en cuenta otros elementos sobre los recursos de las escuelas y sobre su forma de operar para poder llegar a conclusiones sobre la calidad de la educación que imparten.

Si un país obtiene mejores resultados en PISA, no se puede inferir sin más que sus escuelas sean más eficaces, pues el aprendizaje comienza antes de asistir al aula y se desarrolla dentro y fuera de ella. La conclusión válida, en cambio, es que el impacto acumulado de las experiencias de aprendizaje, desde la infancia hasta los 15 años, dentro y fuera de la escuela, trae consigo mejores resultados en los aspectos evaluados por PISA.

Las interpretaciones de los resultados de PISA que centran la atención en el lugar ocupado en ordenamientos simples son relativas y no informan sobre la situación de un país en términos absolutos; en cambio, los porcentajes de jóvenes que se ubican en cada nivel de desempeño sí arrojan un diagnóstico más acertado, como se explicará a continuación.

LOS NIVELES DE DESEMPEÑO

A partir de las distribuciones de los resultados de los alumnos, para cada área o escala medida por las pruebas PISA se definen varias categorías, denominadas de manera genérica *niveles de desempeño*. Cada área tiene distintos niveles de desempeño: en las competencias científica y matemática se establecen seis, y en la lectora, siete. Cada uno de los niveles se define por un rango de puntuaciones. En la parte inferior del Nivel I se añade un nivel más, denominado Debajo del nivel I o Debajo del nivel Ib.

La tabla siguiente presenta, de manera muy general, los niveles de desempeño de las escalas de PISA.

Niveles	Descripción genérica
Nivel 6	Situarse en uno de estos niveles significa que un estudiante tiene potencial para realizar actividades de alta complejidad cognitiva: matemática, científica u otras.
Nivel 5	
Nivel 4	
Nivel 3	Por arriba del mínimo necesario y, por ello, bastante bueno, aunque no el deseable para realizar actividades cognitivas más complejas.
Nivel 2	Identifica el mínimo adecuado para desempeñarse en la sociedad contemporánea.
Nivel 1a	Insuficientes o bajos (en especial el Debajo del nivel I o Debajo del nivel Ib) para acceder a estudios superiores y desarrollar las actividades que exige la vida en la sociedad del conocimiento.
Nivel 1b	
Debajo del nivel I o Ib	

La descripción precisa de los niveles de desempeño es distinta para cada una de las escalas en función de los aspectos conceptuales de las áreas o competencias.

No obstante, para todas las escalas, el Nivel 2 representa el mínimo necesario para que un estudiante pueda seguir aprendiendo en los niveles educativos superiores, o pueda insertarse con éxito en el mercado laboral. Por ello debe preocupar se presenten proporciones considerables de estudiantes ubicados en los niveles bajos.

También es necesario prestar atención a la proporción de jóvenes que logre ubicarse en los niveles más altos de desempeño. Sin embargo, no hay que olvidar que ni los países con mejores resultados consiguen situar en estos niveles a la mayor parte de sus estudiantes de 15 años.

No obstante, si el porcentaje de estudiantes en los niveles más altos es demasiado bajo, el país evaluado no podrá aspirar a formar un número suficiente de especialistas de alto nivel en los diferentes campos disciplinarios y laborales, a menos que se actúe para resolver el problema. Esta situación será un obstáculo importante para el desarrollo de una sociedad avanzada y para una economía competitiva en el mundo globalizado del siglo XXI.

VENTAJAS DE UTILIZAR LOS NIVELES DE DESEMPEÑO

Publicar los resultados por medio de gráficas que indiquen el porcentaje de los estudiantes de un país ubicados en los niveles de desempeño de cada escala, es un tipo de presentación de resultados que permite interpretaciones más ricas que los *rankings*.

Si muchos países consiguen que una mayor proporción de sus jóvenes tenga buenos niveles de desempeño, importará poco que alguno esté en un lugar ligeramente más alto o más bajo en un grupo, pues las distancias que separan a sus integrantes serían pequeñas, probablemente poco significativas. De manera análoga, poca importancia deberá darse al hecho de estar arriba de otros países, si una proporción considerable de los jóvenes del propio sistema educativo no alcanza los niveles mínimos necesarios para una vida adulta plena.

Como se ha explicado antes, los niveles de desempeño de las escalas de PISA se definen de tal manera que el Nivel 2 corresponde al mínimo necesario para que una persona pueda desenvolverse adecuadamente en la vida adulta, mientras los niveles más altos son necesarios para la formación de especialistas de alto nivel.

Tener proporciones considerables de estudiantes por debajo del Nivel 2 debe considerarse como motivo de preocupación, independientemente del lugar que se ocupe en un ordenamiento, y también deberá ser motivo de alerta tener muy pocos alumnos en los niveles de competencia más altos.

PARA FINALIZAR: TRES REGLAS PARA LA INTERPRETACIÓN DE LOS RESULTADOS

Las consideraciones anteriores pueden sintetizarse en la forma de las tres reglas siguientes, cuya aplicación permitirá hacer mejores interpretaciones:

Regla 1

Los ordenamientos simples de promedios (listas de posiciones o *rankings*) son indicadores insuficientes de la calidad educativa.

Regla 2

Los porcentajes de estudiantes en los niveles de desempeño son mejores indicadores del grado en que un país consigue que sus jóvenes desarrollen las competencias necesarias para la vida en las sociedades contemporáneas.

Regla 3

Los ordenamientos de promedios y los porcentajes de estudiantes en ciertos niveles de desempeño adquieren sentido teniendo en cuenta los contextos socioeconómico y cultural de los hogares, así como los de las escuelas.



**Instituto Nacional para la
Evaluación de la Educación**

Barranca del Muerto 341,
Col. San José Insurgentes,
Deleg. Benito Juárez,
C.P. 03900, Ciudad de México.



Comuníquese
con nosotros



Visite nuestro portal