

PISA en detalle

Características del estudio PISA 2003

Primera parte

PISA (*Programme for International Student Assessment*, Programa para la Evaluación Internacional de Alumnos) es un estudio internacional de evaluación educativa de los conocimientos y destrezas de los alumnos a la edad de 15 años, impulsado por la Organización para la Cooperación y el Desarrollo Económico (OCDE). Sus rasgos más importantes son:

- Es un estudio de evaluación comparada a gran escala, en el que los alumnos de decenas de países se enfrentan a los mismos tipos de prueba en condiciones lo más similares posible.
- Es un estudio cíclico que se repite cada tres años con el fin de apreciar la evolución en el tiempo del rendimiento de los alumnos en las materias evaluadas. La primera aplicación de un estudio PISA fue hecha en el año 2000; la segunda, cuyas características comentamos, en 2003 y acaba de terminar la recogida de datos del estudio PISA 2006, cuyos resultados se harán públicos a finales de 2007.
- Es una evaluación de los conocimientos y las destrezas acumuladas por los alumnos a lo largo de su escolaridad hasta la edad de 15 años. No trata de evaluar a los alumnos escolarizados en un determinado curso académico sino a aquellos que han alcanzado una determinada edad biológica.
- Es un estudio orientado a la toma de decisiones para la conducción del sistema educativo (nivel macro). No es un estudio orientado a la investigación académica o al estudio de los procesos educativos de aula o de centro (nivel micro).

PISA evalúa siempre tres materias: Lectura, Matemáticas y Ciencias. Todas las materias son evaluadas en cada ciclo de PISA pero una de ellas ocupa la mayor parte del tiempo de evaluación mientras que de las demás sólo se hace un sucinto sondeo. No se excluye que en algún momento aparezca, de modo esporádico y transversal alguna otra materia. Por ejemplo, en el ciclo 2003 de PISA se introdujo la Solución de problemas como materia transversal.

El estudio también recoge importante información de contexto de los propios alumnos y de los directores de los centros educativos. Esta información contextual, puesta en relación con los resultados de las pruebas de rendimiento, ofrece una rica comparación de los rasgos que caracterizan a los sistemas educativos de los países participantes.

Diseño del estudio

PISA 2003 ha sido llevado a cabo bajo el impulso conjunto de la OCDE y de los países participantes. En el seno de la Dirección para la Educación de la OCDE se ha consti-

tuido un Consejo de Gobierno de PISA (PGB, *PISA Governing Board*) que, junto con la División de Indicadores y Análisis de la mencionada Dirección —que ha ejercido como secretariado permanente— ha ido tomando las decisiones más importantes y

La realización del estudio ha sido confiada a un Consorcio internacional de instituciones de investigación.

de mayor alcance. La realización concreta del estudio –elaboración de las pruebas y los cuestionarios, verificación de su traducción, selección de la muestra, organización del trabajo de campo, diseño de la corrección de las preguntas abiertas y de la entrada y limpieza de datos, realización del análisis y de los cálculos finales– no fue hecha directamente por la OCDE sino por un Consorcio internacional de instituciones de investigación, seleccionado tras un concurso con concurrencia de ofertas. La OCDE se reservó la redacción del informe final y las actividades de difusión de los resultados.

El Consorcio internacional ha sido el mismo que el que realizó la primera aplicación de PISA en el año 2000 y ha estado

formado por las siguientes instituciones: ACER, de Australia; Citogroep, de Holanda; NIER, de Japón; Westat, de EEUU y ETS, de EEUU. El Consorcio movilizó a su vez a expertos individuales y a grupos de expertos para atender a las diversas fases del diseño. Además, y en estrecha colaboración con el Consorcio, en cada país se constituyó uno o varios comités técnicos bajo la dirección de un Coordinador Nacional del Proyecto (NPM, *National Project Manager*).

En España el peso principal del estudio lo llevó un comité técnico formado por un representante de cada una de las 17 Comunidades Autónomas junto con personal del INECSE, antigua denominación del actual Instituto de Evaluación.

El reto de la comparabilidad

En PISA han participado hasta ahora 51 países. En el año 2000 fueron 28 países miembros de la OCDE y 4 países no miembros. En 2002 se repitieron las pruebas de PISA 2000 para 13 países no miembros de la OCDE, en lo que se conoció como PISA Plus. En PISA 2003 los países participantes han sido los 30 miembros de la OCDE y 11 países no miembros.

Tabla 1 Países participantes en PISA

	PISA 2000	PISA Plus 2002	PISA 2003		PISA 2000	PISA Plus 2002	PISA 2003
Albania*		•		Italia	•		•
Alemania	•		•	Japón	•		•
Argentina*		•		Letonia*	•		•
Australia	•		•	Liechtenstein*	•		•
Austria	•		•	Lituania*		•	
Bélgica	•		•	Luxemburgo	•		•
Brasil*	•		•	Macao-China*			•
Bulgaria*		•		Macedonia*		•	
Canadá	•		•	México	•		•
Chile*		•		Noruega	•		•
Taipei-China*		•		Nueva Zelanda	•		•
Corea	•		•	Perú*		•	
Dinamarca	•		•	Polonia	•		•
Eslovaquia			•	Portugal	•		•
España	•		•	Reino Unido	•		•
Estados Unidos	•		•	República Checa	•		•
Finlandia	•		•	Rumania*		•	
Francia	•		•	Rusia*	•		•
Grecia	•		•	Serbia*			•
Holanda	•		•	Suecia	•		•
Hong Kong-China*		•	•	Suiza	•		•
Hungría	•		•	Tailandia*		•	•
Indonesia*		•	•	Túnez*			•
Irlanda	•		•	Turquía			•
Islandia	•		•	Uruguay*			•
Israel*		•					

* Los países con asterisco no son miembros de la OCDE.

La Tabla 1 y el mapa del Gráfico 1, muestran los países que han tenido alguna participación en las evaluaciones PISA realizadas hasta el momento.

El principal reto de diseño que tiene PISA es el de hacer posible la comparabilidad de los datos recogidos dadas las diferencias de idioma, de cultura y de organización escolar.

El Gráfico 1, presenta los países que han participado en los distintos ciclos de PISA

El principal reto de diseño que tiene PISA es el de hacer posible la comparabilidad de los datos recogidos dadas las diferencias de idioma, de cultura y de organización escolar que existen entre los países participantes. PISA se beneficia de la experiencia adquirida por la OCDE en la puesta en marcha de otros proyectos de comparación internacional, como el de producción de los indicadores educativos –proyecto INES–, y por la experiencia de otros estudios anteriores realizados por otras instituciones, muy especialmente del estudio TIMSS de la IEA¹.

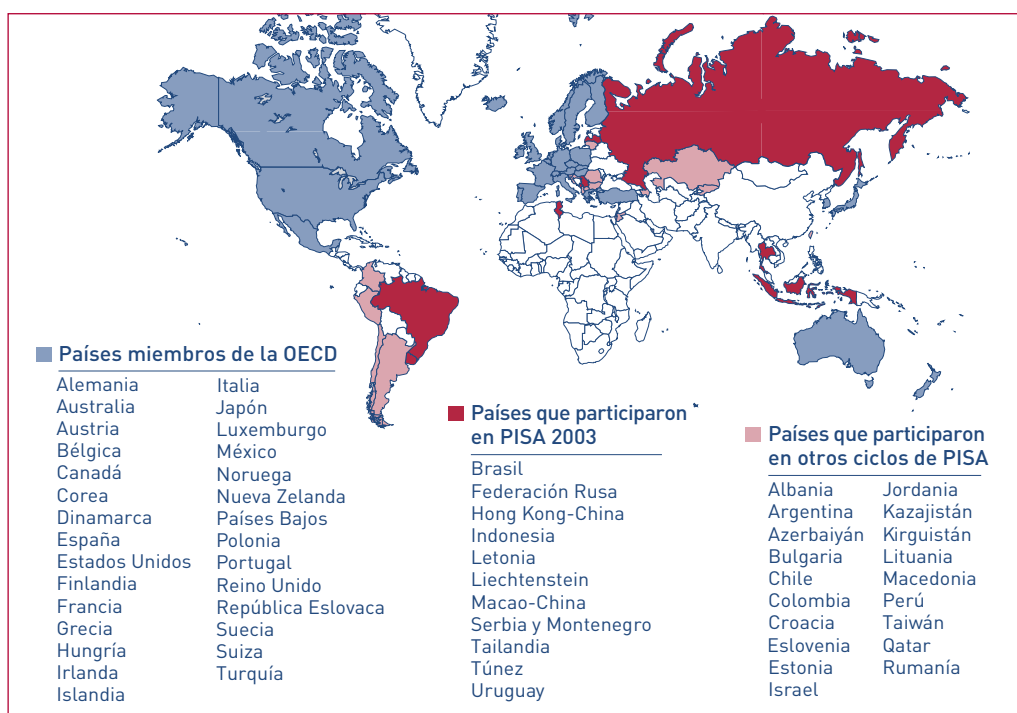
Los procesos de muestreo, de trabajo de campo, de corrección de preguntas abiertas y de grabación y limpieza de los datos no deben introducir a su vez sesgos diferenciales entre países. Para evitarlo, el Consorcio generó manuales específicos con instrucciones precisas y pormenorizadas. A lo largo de todos los pasos en la preparación y realización del trabajo de

Para garantizar unos niveles aceptables de comparabilidad ha sido necesario, por un lado, abordar la traducción de los instrumentos de evaluación y garantizar que su contenido fuera lingüísticamente equivalente. Los instrumentos han sido elaborados en inglés y francés (idiomas oficiales de la OCDE), teniendo ambos idiomas el carácter de versiones fuente. Posteriormente cada país procedió a traducir o a adaptar el contenido de estos instrumentos.

Las traducciones y adaptaciones fueron luego verificadas por una empresa independiente radicada en Bruselas.

campo, el Consorcio ha hecho un seguimiento detallado del cumplimiento de los procedimientos establecidos mediante formularios que cada país debía cumplimentar o mediante frecuentes comunicaciones por correo electrónico. Los Coordinadores Nacionales de Proyecto fueron convocados a dos reuniones anuales durante los tres años de desarrollo de este ciclo de PISA.

Gráfico 1 Países que han participado en los distintos ciclos de PISA



¹ TIMSS es el Estudio internacional de tendencias en Matemáticas y Ciencias (Trends in International Mathematics and Science Study) que con carácter cíclico pone en marcha la IEA (International Association for the Evaluation of Educational Achievement, Asociación internacional para la evaluación del rendimiento educativo), con sede en Ámsterdam.

Todas las escalas de PISA se construyen de modo que la media de los países de la OCDE tenga un valor de 500 puntos y una desviación típica de 100.

Por último, es necesario también garantizar que los resultados se expresen en una métrica que permita la comparación tanto en el espacio –entre países– como en el tiempo –entre ciclos PISA–. Los resultados que miden en rendimiento de los alumnos se han calculado según un modelo TRI (Teoría de Respuesta al Ítem) y se han expresado mediante escalas cen-

tradas sobre la media de los países pertenecientes a la OCDE e invariante entre los distintos ciclos de PISA, de modo que se puedan apreciar con precisión las posibles tendencias en el tiempo. Todas las escalas de PISA se construyen de modo que la media de los países de la OCDE tenga un valor de 500 puntos y una desviación típica de 100.

Enfoque no curricular y no académico

La dificultad de encontrar suficientes puntos comunes entre los sistemas educativos de los países participantes hizo que desde sus inicios PISA evitara evaluar los conocimientos de los alumnos que se encuentran en un determinado curso del sistema educativo, y se dirigiera a evaluar los de una determinada edad biológica. Se ha elegido la edad de 15 años como aquella que, en la mayoría de los países, está razonablemente cercana a la finalización de la escolaridad obligatoria y al momento en el que alumno debe tomar la decisión de incorporarse al mundo laboral o de proseguir estudios más avanzados.

En España, y a diferencia de muchos otros países, los alumnos de 15 años no están dispersos en diferentes programas de estudios sino que todos ellos están cursando la Educación Secundaria Obligatoria (ESO). La mayor parte de estos alumnos cursan el cuarto curso, pero también se encuentran alumnos de esa edad en tercero y segundo, debido verosímelmente a haber repetido algún curso en su escolaridad anterior.

PISA no indaga lo que los alumnos saben sino lo que los alumnos saben aplicar.

La opción por una edad biológica en lugar de un curso académico ha facilitado también un enfoque no academicista en la orientación de las pruebas. En lugar de comprobar si los alumnos han aprendido lo que se les ha enseñado, PISA adopta el enfoque de comprobar si los alumnos están suficientemente equipados en conocimientos y destrezas para integrarse en la vida adulta. PISA indaga si la formación adquirida por los alumnos a lo largo de su recorrido escolar les permite enfrentarse a los retos de la vida real como ciudadanos activos y responsables. PISA no comprueba, por tanto, los conocimientos en sí mismos sino que busca averiguar si los alumnos son capaces de activar esos conocimientos en contextos de vida cotidiana. En suma, PISA no indaga lo que los alumnos saben sino lo que los alumnos saben aplicar.

El nombre que se le da a esta formación en inglés es el de *literacy*, término tradicionalmente asociado en español a alfabetización, pero que lo desborda en su uso más literal. PISA no es un estudio de alfabetización en el sentido del manejo del alfabeto sino en un sentido más global y comprensivo de manejo de los recursos necesarios para la vida adulta. Lo importante es que el referente no está en los estudios posteriores ni en una formación destinada en último término al progreso de la ciencia, sino en la posesión de un equilibrado conjunto de herramientas adaptado a la vida futura de un ciudadano responsable. Las pruebas de evaluación de PISA se construyen atendiendo a un contexto adulto y cercano a lo cotidiano, con una orientación denominada *auténtica* y con un contenido similar a lo que se consideran competencias básicas, como opuestas a competencias académicas.

Carácter cíclico y rotatorio del estudio

PISA es un estudio que se repite cada tres años y que examina los conocimientos y destrezas en tres materias –Lectura, Matemáticas y Ciencias–. Estas tres materias suelen cursarse en todos los sistemas educativos –aunque en el caso de las Ciencias con nombres y contenidos dispares– y, además, su currículo suele ser más neutro y menos dependiente del entorno cultural específico de cada país.

PISA examina siempre Lectura, Matemáticas y Ciencias, pero una de ellas con especial intensidad en cada ciclo.

Todas estas tres materias son evaluadas en cada ciclo, pero no con la misma intensidad. Una de ellas ejerce de materia principal del ciclo y a ella se le dedica aproximadamente las dos terceras partes del tiempo de evaluación. Las otras ejercen de materias

secundarias y entre ambas se reparten el tiempo restante. La materia principal rota entre los distintos ciclos de PISA: en PISA 2000 fue la Lectura, en PISA 2003 han sido las Matemáticas y en PISA 2006 son las Ciencias.

En el ciclo 2003 de PISA apareció, además, una cuarta materia de carácter transversal: la Solución de problemas. Fue incluida a título de ensayo y a la vista de los resultados recogidos (la correlación con Matemáticas fue muy alta) se ha decidido no volver a incluir esta materia en

futuros ciclos de PISA, haciendo públicas todas las preguntas utilizadas, sin reservar ninguna para futuras aplicaciones.

El reparto del tiempo de evaluación en PISA 2003 quedó como sigue: Matemáticas, 55%; Lectura, Ciencias y Solución de problemas, 15% para cada una.

Los instrumentos de evaluación

Al igual que en año 2000, la evaluación PISA 2003 consistió en la aplicación a los alumnos de una prueba de dos horas de duración, con un pequeño descanso a la mitad, y posteriormente de un cuestionario de contexto de unos 30 minutos. Al mismo tiempo se le pidió al director o a un miembro del equipo directivo que respondiera a un cuestionario sobre el centro educativo de 20 minutos de duración. En España no se aplicaron otros dos componentes opcionales: un cuestionario sobre conocimientos informáticos y otro sobre expectativas de estudio.

Preguntas, unidades, bloques y cuadernos de prueba

PISA evalúa los conocimientos y destrezas mediante pruebas de papel y lápiz. Las pruebas constan de preguntas en varios formatos. Las preguntas suelen ser cerradas o de respuesta múltiple sencilla (cuatro o cinco alternativas), de respuesta múltiple compleja (conteniendo varias alternativas dicotómicas del tipo “Sí/No” o “Verdadero/Falso”), abiertas de respuesta corta y abiertas de respuesta larga o construida. En estas últimas el alumno tiene que redactar con-

testaciones de más de una línea o mostrar los cálculos que le han llevado a la solución.

Aunque en las pruebas de PISA las preguntas son las unidades mínimas de puntuación, se suelen presentar agrupadas en unidades. Una unidad consiste típicamente en un estímulo común, sea este un relato o la descripción de una situación problemática, muy frecuentemente con apoyo gráfico, al que le siguen unas pocas preguntas relacionadas con dicho estímulo común. Las

Las preguntas están agrupadas en unidades, con textos y gráficos introductorios comunes.

unidades de una misma materia a su vez se agrupan en bloques de media hora y cuatro de estos bloques forman un cuaderno de prueba. En PISA 2003 hubo trece cuadernos de prueba distintos en los que cada bloque se encontraba repetido en posición inicial, media anterior, media posterior y final en cuatro cuadernos.

Al finalizar un ciclo PISA, una parte de las

unidades utilizadas se liberan o se hacen públicas, reservándose el resto para futuras aplicaciones. Estas preguntas reservadas son esenciales para asegurar que las puntuaciones otorgadas en cada ciclo se expresan en una escala común y, por tanto, para que pueda darse una comparación rigurosa en el tiempo. El Instituto de Evaluación ha publicado en sus páginas web todas las preguntas liberadas por PISA hasta el momento².

Los formatos de las preguntas suelen ser: cerradas o de respuesta múltiple sencilla, de respuesta múltiple compleja, abiertas de respuesta corta y abiertas de respuesta larga.

Marcos teóricos de referencia

La elaboración de las unidades ha sido el fruto de un trabajo de cooperación internacional. En primer lugar los grupos de expertos internacionales han elaborado un marco teórico de referencia para cada materia, donde se especifican los objetivos de evaluación y las competencias que se deseaba medir.

En la segunda parte de este artículo se presentan, en cuatro recuadros diferentes, los resúmenes de los marcos teóricos de las cuatro materias evaluadas en PISA 2003. Los marcos teóricos también han especificado los porcentajes deseados según el tipo de respuesta solicitada (múltiple, construida cerrada y construida abierta)³. A la vista de estos marcos teóricos los distintos países propusieron unidades que fueron estudiadas por los grupos de expertos y luego revisadas en cada país. Las unidades fueron sometidas a un ensayo de campo en una amplia prueba piloto realizada en 2002. Esta prueba piloto permitió medir las características psicométricas de las preguntas y realizar la selección de las que formarían parte de la prueba final aplicada al año siguiente.

En PISA 2003 hubo bloques de preguntas suficientes para seis horas y media de examen, pero a cada alumno sólo se le pidió un esfuerzo de dos horas, por lo que no todos los alumnos contestaron a las mismas preguntas. La mayor abundancia en material de evaluación redundó en una mejor exploración de la materia examinada.

Haciendo uso de la técnica denominada *muestreo matricial* se produjeron trece cuadernos de prueba distintos, conteniendo cada uno cuatro bloques. Todos los bloques aparecían en posición inicial, final e inter-

media dentro de algún cuaderno, ya que la ubicación de las preguntas en una prueba afecta a su dificultad percibida: las situadas al final de un cuaderno resultan más difíciles que las situadas al comienzo. Algunos bloques también contenían unidades ya utilizadas en PISA 2000 y que tenían la función de facilitar la equiparación de las puntuaciones del ciclo 2003 con el ciclo 2000. En cada centro se utilizaron los trece cuadernos mediante una asignación aleatoria y rotatoria, de modo que los alumnos que se sentaban juntos no estaban contestando al mismo cuaderno.

² Las preguntas liberadas se encuentran en las siguientes cuatro publicaciones del Instituto de Evaluación:

- Preguntas planteadas en PISA 2000. Lectura, Matemáticas y Ciencias (pisa2000liberadas.pdf);
- PISA 2003: Preguntas liberadas. Matemáticas y Solución de problemas (pisa2003liberados.pdf);
- Programa PISA. Ejemplos de ítems de Conocimiento Científico (itemscienciaspisa.pdf);
- Programa PISA. Pruebas de Comprensión Lectora (pisa2000cuadlectura3.pdf);

todas ellas disponibles en la página web del Instituto anteponiendo a los nombres de fichero PDF que se mencionan entre paréntesis la especificación “<http://www.ince.mec.es/pub/>”.

³ Puede consultarse en <http://www.ince.mec.es/pub/marcoteoricopisa2003.pdf> la traducción al castellano de los mencionados marcos teóricos.

Selección de la muestra de centros y de alumnos

Las especificaciones internacionales de muestreo exigían un mínimo de 4.500 alumnos y de 150 centros docentes por país. El muestreo podía hacerse proporcional a un cierto número de estratos, y el plan de muestreo de cada país debía ser negociado previamente con el Consorcio internacional. Al igual que en PISA 2000, España eligió realizar una muestra estratificada y bi-etápica. Los estratos fueron 34: las 17 Comunidades Autónomas y, dentro de ellas, la titularidad de los centros docentes (públicos o privados). Las etapas fueron dos: selección inicial de centros y selección posterior de alumnos dentro de esos centros.

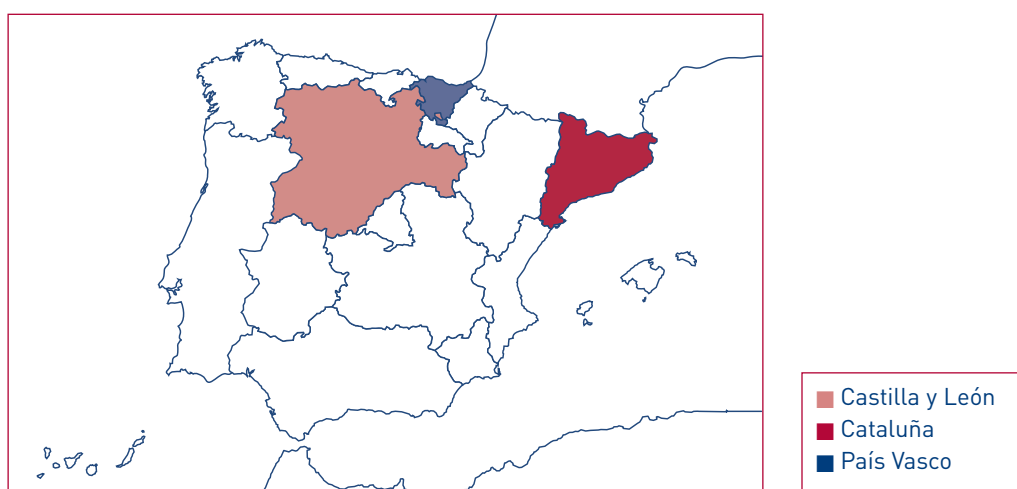
Al aumentar su muestra, los territorios consiguen suficiente precisión estadística para entrar en la comparación PISA.

En la primera etapa se seleccionó al azar una muestra de centros de educación secundaria proporcional a la población escolar matriculada en cada uno de los 34 estratos. En la segunda etapa se seleccionaron al azar un máximo de 35 alumnos de 15 años dentro de cada centro, con independencia del curso en el que estuvieran matriculados. Es decir, los alumnos retrasados o repetidores así como los adelantados fueron objeto de evaluación si tenían 15 años en el momento de la administración de las pruebas.

Tres Comunidades Autónomas –Castilla y León, Cataluña y País Vasco– decidieron ampliar la muestra correspondiente a su territorio de modo que fuera factible obtener una desagregación de sus resultados con suficiente precisión estadística como para poder compararlos con los de otros territorios y países participantes en PISA. En el caso de estos territorios, denominados “subnacionales” en la literatura de PISA, los requerimientos de la muestra ampliada eran un mínimo de 50 centros docentes y 1.500 alumnos.

Gráfico 2 Comunidades autónomas que ampliaron su muestra en PISA 2003

El Gráfico 2 presenta las CC.AA que ampliaron su muestra en PISA 2003.



Al elaborar la muestra de centros, se excluyeron los específicos de educación especial. Al elaborar la muestra de alumnos en cada centro, fueron excluidos los alumnos con discapacidades que les impidiesen rellenar los cuadernillos de la prueba. También fueron excluidos aquellos alumnos que por no tener un año de escolaridad completa en España no dominaban suficientemente el idioma de la prueba.

Tabla 2 Muestra y población representada en PISA 2003

	España	Castilla y León	Cataluña	País Vasco	Método de cálculo	Descripción
a	454.064	24.210	62.946	18.160		Población total de 15 años
b	418.005	21.580	61.829	17.753		Población total de 15 años escolarizada
c	418.005	21.580	61.829	17.753		Población objetivo
d	1.639	109	576	15		Población excluida al excluir centros de educación especial específica
e	416.366	21.471	61.253	17.738	c-d	Población total tras excluir centros completos y antes de excluir alumnos en centros que participan
f	0,39	0,51	0,93	0,08	d/c·100	Porcentaje de exclusión de alumnos al excluir centros
g	10.791	1.490	1.516	3.885		Número de alumnos que participan
h	344.372	18.224	50.484	16.978		Número ponderado de alumnos que participan
i	591	95	61	56		Número de alumnos excluidos en centros que participan
j	25.619	1.057	1.847	252		Número ponderado de alumnos excluidos en centros que participan
k	6,92	5,48	3,53	1,46	j/(h+j)	Porcentaje de exclusión de alumnos en centros que participan
l	7,29	5,96	4,43	1,55	$\frac{(f/100)+(k/100)}{1-(f/100)} \cdot 100$	Porcentaje de exclusión de alumnos total

La Tabla 2 presenta los principales datos acerca de la muestra y de cómo representa a la población de jóvenes de 15 años.

La muestra efectiva, una vez realizadas las sustituciones pertinentes y descontados los excluidos, abandonos y casos perdidos, asciende a las cifras siguientes:

Tabla 3 Muestra efectiva de centros y alumnos en los distintos territorios españoles

	Centros			Alumnos		
	Públicos	Privados	Total	Públicos	Privados	Total
España	203	180	383	5.606	5.185	10.791
de los cuales:						
Castilla y León	30	21	51	878	612	1.490
Cataluña	27	23	50	810	706	1.516
País Vasco	58	83	141	1.502	2.383	3.885

La Tabla 3 presenta los datos acerca de la muestra efectiva de centros y alumnos en los distintos territorios españoles.

El País Vasco, en particular, deseó introducir un tercer nivel de estratificación: los tres modelos lingüísticos en los que se imparte la enseñanza secundaria en esa comunidad autónoma (A: todo en castellano; B: en ambos idiomas y D: todo en euskera), lo que provocó un sobremuestreo considerable en el número de centros y de alumnos⁴ de este territorio. La muestra del País Vasco se desglosó por modelos lingüísticos del siguiente modo:

Tabla 4 Muestra de centros y alumnos por modelo lingüístico en el País Vasco

	Centros			Alumnos		
	Públicos	Privados	Total	Públicos	Privados	Total
Modelo A	11	35	46	256	1.019	1.275
Modelo B	16	24	40	365	636	1.001
Modelo D	31	24	55	881	728	1.609
Total País Vasco	58	83	141	1.502	2.383	3.885

La Tabla 4 presenta la muestra de centros y alumnos por modelo lingüístico en el País Vasco.

⁴ Este sobremuestreo del País Vasco no produce ningún sesgo en los resultados del conjunto del Estado al quedar perfectamente controlado por la variable de ponderación pertinente.

Aplicación de los instrumentos, codificación y grabación de los datos

Trabajo de campo

La aplicación de las pruebas y cuestionarios fue realizada en el mes de mayo de 2003 por una empresa especializada que se encargó de todas las tareas logísticas y organizativas del trabajo de campo.

Los instrumentos fueron aplicados en catalán en todos los centros de Cataluña e Islas Baleares. En Galicia y en la Comunidad Valenciana los alumnos podían escoger, en el momento de la prueba, la lengua –gallego o valenciano, respectivamente– en la que

deseaban contestar. En el País Vasco las autoridades educativas decidieron la lengua en la que debía contestar cada alumno, atendiendo sobre todo a la lengua hablada habitualmente en el hogar familiar, no coincidente en bastantes casos con la lengua estipulada por el modelo lingüístico en el que el alumno estaba matriculado. En Navarra, sin embargo, fue la modalidad lingüística del centro la que decidió la lengua en la que el alumno debería contestar la prueba.

La Tabla 5 resume, en términos porcentuales, la utilización de las diversas lenguas en la aplicación de las pruebas de PISA 2003 en España.

Tabla 5 Lengua de aplicación en España de las pruebas PISA 2003

CCAA	Castellano	Catalán	Gallego	Valenciano	Euskera	Total
Baleares		100%				100%
Cataluña		100%				100%
Galicia	39%		61%			100%
Navarra	76%				24%	100%
País Vasco	85%				15%	100%
C. Valenciana	86%			14%		100%
Resto del estado	100%					100%
TOTAL	77,4%	15,1%	1,6%	0,4%	5,5%	100%

Corrección y codificación de las preguntas abiertas

Tras la aplicación de las pruebas hubo que proceder a corregir y puntuar manualmente las respuestas dadas a las preguntas abiertas, tarea especialmente delicada en una evaluación de carácter comparativo internacional en la que los sesgos culturales están siempre al acecho.

Esta tarea fue encomendada a unos correctores especialmente entrenados, todos ellos profesores de secundaria en ejercicio, que contaban con una extensa guía internacional con criterios de corrección específica-

dos de modo pormenorizado y preciso. Además, y para tener una medida del posible sesgo introducido por la intervención de correctores humanos, un subconjunto de los cuadernillos fue corregido por cuatro correctores distintos para poder comparar sus puntuaciones y asegurarse de que los criterios de corrección fueron aplicados de un modo consistente.

En el análisis posterior no se detectó ningún problema de fiabilidad importante en la corrección practicada en España.

Las preguntas abiertas fueron corregidas manualmente, tarea especialmente delicada en una evaluación internacional

Grabación, depuración y construcción de la base de datos

El cálculo de las puntuaciones parte de modelos TRI (Teoría de Respuesta al Ítem) y utiliza valores plausibles.

Una vez corregidas las preguntas abiertas, los cuadernillos de prueba y los cuestionarios fueron grabados en soporte informático por una casa especializada y el conjunto de ficheros resultante fue sometido a un exhaustivo control y depuración en dos fases: una realizada en España y una segunda realizada por el Consorcio técnico internacional en su sede de Australia.

Con los ficheros depurados de todos los países, el Consorcio procedió a calcular los parámetros de los ítems y asignar las puntuaciones obtenidas por cada alumno en las distintas materias, a través de un modelo de respuesta al ítem y utilizando la metodología de los valores plausibles.

También estableció las variables de ponderación tanto para los datos de los alumnos como para los de los centros, así como para el procedimiento de replicación repetida equilibrada⁵ utilizado para calcular los errores típicos y los intervalos de confianza.

Así mismo sintetizó en una serie de índices y constructos estadísticos una buena parte de las respuestas de los alumnos y de los directores de los centros a sus respectivos cuestionarios de contexto destacando, en particular, el índice ESEC de estatus socioeconómico y cultural. Por último, puso a disposición del Secretariado de la OCDE y los países participantes la base de datos resultante.

Escalas y niveles de rendimiento

Escalas de rendimiento

Las escalas de rendimiento que contienen las puntuaciones de los alumnos son variables tipificadas pero transformadas a una escala de media 500 y desviación típica 100, lo que hace más fácil su lectura y permite prescindir en la práctica de los decimales. Las escalas se construyen de modo

que el valor 500 corresponda a la media de los países miembros de la OCDE (cada país contribuye con el mismo peso). De un modo similar, es esperable que, globalmente, dos tercios de la población de alumnos obtengan una puntuación entre 400 y 600 puntos.

En principio, las escalas de rendimiento, una vez construidas, son invariantes entre distintos ciclos PISA, gracias a que las pruebas incluyen preguntas puente, comunes a más de un ciclo, que permiten equiparar las escalas de puntuaciones y situarlas en una misma métrica. De este modo, una puntuación de 500 puntos obtenida en PISA 2003 significa que se ha logrado el mismo rendimiento que el que fue promedio de la OCDE en PISA 2000.

Pero se han detectado problemas de equiparación cuando se intentan incorporar puntuaciones obtenidas de pruebas con muchas preguntas a escalas estaban construidas a partir de pruebas de pocas preguntas, como es el caso de las materias secundarias en PISA. Por eso se ha tomado la decisión de no considerar invariante la escala hasta que la materia correspondiente haya sido la materia principal de un ciclo PISA, y esa es la razón por la que la escala de Matemáticas, que ya se había construido en PISA 2000 cuando las Matemáticas eran una

Las escalas de rendimiento en PISA tienen una media de 500 puntos y una desviación típica de 100. La media corresponde al conjunto de países de la OCDE.

⁵ *Balanced repeated replication*. PISA utilizó la variante de Fay, con un factor deflactor de valor 0,5. La base de datos internacional de PISA 2003 contiene 80 replicados, gracias a los cuales es posible, mediante software especializado o mediante el procedimiento detallado en el manual de análisis de datos (PISA 2003 Data Analysis Manual, Paris, OCDE, 2005 –hay una versión traducida a castellano en <http://www.ince.mec.es/pub/pisamanualdatos.pdf>), hallar el error típico, y su consiguiente intervalo de confianza, de los estadísticos que se calculen a partir de dicha base de datos.

materia secundaria, se ha vuelto a construir de nuevo en PISA 2003, siendo esta última la considerada como escala invariante para el futuro.

Los resultados obtenidos en Matemáticas en PISA 2000 se han reescalado de acuerdo a la nueva escala, para permitir una comparación rigurosa con los resultados de PISA 2003. Del mismo modo en PISA 2006, ciclo en el que las Ciencias serán la materia principal por primera vez, se volverá a calcular de nuevo la escala de Ciencias y a partir de ella se recalcularán las puntuaciones obtenidas en los ciclos anteriores PISA 2000 y PISA 2003, en donde las Ciencias han sido una materia secundaria.

Niveles de rendimiento

A cada nivel de rendimiento se le asocia una descripción de las competencias y capacidades típicas del alumno de ese nivel.

En una evaluación PISA, la misma escala de puntuaciones sirve tanto para estimar las capacidades de los alumnos como para caracterizar la dificultad de las preguntas. En los alumnos, su puntuación en una prueba indica el nivel de dificultad de las preguntas que son típicamente capaces de responder. En las preguntas, la puntuación indica su dificultad, es decir la probabilidad típica de ser contestadas correctamente por los alumnos.

Para dar un mayor sentido educativo a la dispersión numérica de las puntuaciones que se producen en una evaluación PISA,

estas se agrupan en varios niveles de rendimiento. Con cada nivel de rendimiento los grupos de expertos asocian una descripción de las competencias y capacidades que típicamente demuestran los alumnos al alcanzar el rango de puntuaciones propio del nivel. En general, se logra una descripción más precisa y más detallada de las capacidades demostradas por los alumnos cuanto más abundan las preguntas relativas a estas capacidades, y esto ocurre cuando el área de conocimiento descrito ha sido materia principal de un ciclo PISA.

En PISA 2000 la Lectura fue la materia principal de evaluación y como resultado del análisis de la dispersión de las puntuaciones obtenidas por los alumnos fue posible distinguir seis niveles de rendimiento. Estos niveles de rendimiento tienen definidos unos límites precisos que son aplicables también a los resultados en Lectura en PISA 2003, ya que la escala en las que se mide el rendimiento de los alumnos es la misma en ambos ciclos. En el presente ciclo PISA 2003 la materia principal ha sido Matemáticas y la dispersión en las puntuaciones han resultado ser algo mayor que en Lectura, con lo que los niveles de rendimiento han sido fijados en siete.

En Lectura la anchura de los niveles de rendimiento es de 72 puntos. En Matemáticas es de 62. En ambas materias se dan dos propiedades estadísticas:

- (a) la probabilidad de que un alumno con una determinada puntuación de rendimiento conteste correctamente a una pregunta con la misma puntuación de dificultad es del 62%
- (b) la probabilidad de que un alumno en el límite inferior de un nivel conteste correctamente a cualquier pregunta de ese mismo nivel de dificultad es de al menos un 50%.

En Ciencias, por no haber sido aún materia principal de un ciclo PISA, no ha habido suficiente número de preguntas para la definición de niveles detallados, por lo que hasta ahora sólo ha sido posible distinguir entre un nivel bajo, medio o alto de rendimiento sin demasiada precisión.

En Solución de problemas el número de preguntas también ha sido escaso y por ello los niveles definidos han sido sólo cuatro.

Una descripción de las capacidades de los alumnos asociadas a los diferentes niveles de rendimiento puede encontrarse, para cada materia, en los cuadros que se encuentran en la segunda parte de este artículo, en el siguiente número de *Apuntes del Instituto de Evaluación*.

Más información sobre este tema
en nuestra página web: www.ince.mec.es

Próximamente
www.institutodeevaluacion.mec.es



Instituto de
evaluación

San Fernando del Jarama, 14
28002 Madrid (España)
info.institutodeevaluacion@mec.es