

El Proyecto PISA: su Aplicación en México

Rafael Santiago Vidal Uribe
Ma. Antonieta Díaz Gutiérrez
Javier de Jesús Loyola del Río

CUADERNO No. 9



Instituto Nacional para la
Evaluación de la Educación

COLECCIÓN CUADERNOS
DE INVESTIGACIÓN

ISSN 1665-9457

El Proyecto PISA: su Aplicación en México

Rafael Santiago Vidal Uribe*
Ma. Antonieta Díaz Gutiérrez
Javier de Jesús Loyola del Río

CUADERNO No. 9

*Los autores pertenecen al Instituto Nacional para la Evaluación de la Educación
Este es un resumen ejecutivo de su estudio, el texto completo puede consultarse en: www.inee.edu.mx

MÉXICO, NOVIEMBRE, 2003

CONTENIDO

▲ Presentación	3
▲ El Proyecto PISA	4
—Aplicaciones de PISA	5
—El Proyecto PISA y otras pruebas internacionales	5
▲ Conocimientos y habilidades para la vida, primeros resultados de PISA 2000	6
—Cómo se miden los conocimientos y aptitudes en PISA	7
—Significado y resultados en los niveles de competencia de lectura	7
—Significado y resultados en los niveles de competencia en matemáticas y ciencias	11
—Sobre la interpretación de los resultados	15
▲ Metodología para el diseño y aplicación de las pruebas PISA	16
—Diseño de instrumentos	16
—Aspectos operativos	17
—Operaciones de campo	18
—Procesamiento de datos	19
—Construcción de escalas	19
▲ PISA 2003 en México	20
—Aplicación	20
—Proceso de codificación	22
▲ Conclusiones	25

PRESENTACIÓN

¿Se encuentran los estudiantes bien preparados para enfrentar los retos del futuro? ¿Son ellos capaces de analizar, razonar y comunicar sus ideas efectivamente? ¿Tienen la capacidad de continuar aprendiendo a lo largo de su vida? Éstas son preguntas para las que los padres de familia, maestros, autoridades educativas, estudiantes y público en general deben tener una respuesta.

El Programa para la Evaluación Internacional de los Estudiantes (PISA por sus siglas en inglés) promovido y organizado por la OCDE, es un esfuerzo de colaboración internacional que representa un compromiso de los gobiernos de los países de la OCDE, para monitorear regularmente los resultados de los sistemas educativos de los países participantes, en términos de logros de los estudiantes. Al proporcionar información relevante sobre los resultados educativos y contextos en los que éstos ocurren, PISA se propone fundamentalmente ayudar a reflexionar a los países participantes, sobre la política y metas educativas. Asimismo, proporcionar datos para el establecimiento de estándares educativos, y ayudar a una mejor comprensión de las causas y consecuencias de las deficiencias que se observen.

PISA hace énfasis en la evaluación de las habilidades que los estudiantes necesitarán en su vida diaria, y por eso su grupo objetivo es la población de 15 años que se encuentra en el momento de ingresar a la educación postsecundaria o a punto de ingresar a la vida del trabajo; por ello, el estudio se enfoca a evaluar si los estudiantes pueden aplicar el conocimiento que han aprendido en la escuela, más que al contenido de los planes de estudio o al currículo de la escuela secundaria.

A la fecha, PISA es el estudio internacional más riguroso para evaluar el desempeño de los estudiantes. Las decisiones acerca de la proyección y naturaleza del mismo y sobre la información que debe ser recabada, están a cargo de expertos de reconocido prestigio internacional. Existe, también un mecanismo estricto de control de calidad para la traducción, muestreo, aplicación, calificación y procesamiento de datos. Todos los pasos desde el diseño, aplicación, análisis, etcétera, son monitoreados regularmente por agentes externos a las instancias responsables de operarlo. Como consecuencia de lo anterior, los resultados

tienen un alto grado de validez y confiabilidad, y pueden ayudar a mejorar significativamente la comprensión de los resultados de la educación en los países que participan en el proyecto.

Es importante mencionar que PISA es un programa diseñado específicamente para impactar la política educativa (*policy-oriented*), y aportar sistemáticamente datos, informes, análisis y reportes dirigidos a los tomadores de decisiones y a la sociedad en general sobre los asuntos más relevantes de la política educativa.

Este informe pretende mostrar los propósitos del proyecto PISA; los resultados más importantes y la manera como este tipo de estudios se desarrolla; así como la participación de México para llevarlo a cabo.

EL PROYECTO PISA

El Programa Internacional para la Evaluación de los Estudiantes (PISA) es un proyecto de los países miembros de la OCDE, el cual tiene como propósito medir qué tan bien están preparados los jóvenes actualmente para enfrentar los retos del mundo moderno.

El proyecto parte del supuesto de que obtener datos sobre el aprovechamiento escolar de los estudiantes, su contexto individual, familiar, y factores institucionales y sociales, es de gran ayuda para explicar las diferencias en su desempeño.

Las principales características del estudio son:

- Orientación hacia la política educativa con diseño y métodos de reporte determinados por las necesidades de los gobiernos de los países participantes, para derivar del estudio lecciones de política educativa.
- Enfoque innovador hacia la competencia no sólo de lectura, sino también de ciencias y matemáticas.
- Enfoque en la demostración de conocimientos y habilidades, de manera que sean relevantes para la vida diaria.
- Amplitud en la cobertura geográfica, con 45 países participando, lo cual representa un tercio de la población mundial.
- Regularidad con un compromiso de repetir el estudio cada tres años.
- Naturaleza colaborativa, en el sentido de que los gobiernos de los países participantes conducen en conjunto el proyecto, junto con un consorcio de instituciones líderes en el mundo en el campo de la evaluación, y con conocimiento científico de punta.

APLICACIONES DE PISA

El Proyecto PISA se realizó por primera vez en el año 2000 con la participación de 32 países, (28 países de la OCDE y cuatro países no miembros). En el año 2002, otros 12 países no miembros de la OCDE llevaron a cabo una aplicación complementaria; sin embargo, tres países (Holanda, Rumania y China) no pudieron estar presentes en el estudio, llegando a un total de 41 países reportados en el documento *Aptitudes Básicas para el mundo de mañana. Otros resultados del proyecto PISA 2000*. A esta aplicación complementaria del primer estudio se le ha denominado coloquialmente como *PISA plus*.

Órganos de coordinación de PISA

Órgano	Función
Consejo de Países Participantes (BPC)	<ul style="list-style-type: none"> • Fijar las prioridades de PISA • Supervisar el programa
Grupos de Expertos de PISA (FEG)	<ul style="list-style-type: none"> • Asegurar la validez y confiabilidad de la prueba
Gerentes Nacionales del Proyecto (NPM)	<ul style="list-style-type: none"> • Implementación del proyecto en cada país
Consortio PISA	<ul style="list-style-type: none"> • Diseño de los instrumentos • Manejo de datos • Calificación de las pruebas
Secretariado de la OCDE	<ul style="list-style-type: none"> • Manejo del proyecto por parte de la OCDE • Interlocutor entre el BPC y el Consortio PISA

Muestra

Se utilizan muestras representativas entre 4 mil 500 y 10 mil estudiantes en cada país, de manera que es posible realizar inferencias para el país en su conjunto, no así para regiones o estados. En el caso particular de México en la aplicación del 2003, se decidió ampliar la muestra hasta más de 30 mil estudiantes, para poder derivar conclusiones a nivel estatal y no sólo a nivel país.

EL PROYECTO PISA Y OTRAS PRUEBAS INTERNACIONALES

El proyecto PISA es diferente a proyectos anteriores de evaluación en lo siguientes aspectos:

Origen. La iniciativa ha sido tomada por los gobiernos participantes y el estudio fue diseñado para servir a su interés de fortalecer los temas de política educativa.

Regularidad. El compromiso de abarcar distintas áreas de evaluación, en detalle cada nueve años y con actualizaciones cada tres, permitirá a los países un seguimiento regular y predecible de sus progresos para alcanzar las principales metas académicas.

Edad del grupo. La evaluación de los jóvenes que están acabando el periodo de escolarización obligatoria, aporta un indicador útil acerca del rendimiento de los sistemas educativos. Aunque la mayor parte de los jóvenes de los países de la OCDE continúan su educación después de los 15 años, esta edad se sitúa generalmente cerca del final del periodo de escolarización obligatoria, en el que los alumnos siguen un currículo básico común, en términos generales. En este punto de la formación, resulta útil determinar el grado en el cual los alumnos han adquirido el conocimiento y las destrezas necesarias, que les ayudarán en el futuro para desenvolverse en la vida, incluyendo las opciones específicas de aprendizaje que pueden elegir más tarde.

Conocimiento y las destrezas evaluadas. Éstas no se definen a partir de un denominador común de los currículos nacionales, sino en términos de las destrezas que se consideran imprescindibles para la vida en el futuro. Ésta es la característica más relevante y novedosa del proyecto PISA. Resultaría arbitrario hacer una distinción exacta entre destrezas *académicas* y destrezas *para la vida*, dado que los centros educativos siempre han buscado la preparación de los jóvenes para la vida, pero esta distinción no deja de ser relevante.

Tradicionalmente, los currículos han sido estructurados en gran medida en bloques de información y dominio de técnicas concretas, de modo que se han centrado menos en las destrezas que se desarrollan en cada área para su aplicación futura en la vida y, menos aún, en las competencias más generales para resolver problemas y aplicar ideas y razonamientos propios en las situaciones cotidianas. El proyecto PISA no excluye los conocimientos y la comprensión basados en el currículo. Sin embargo, los evalúa, sobre todo, en términos de adquisición de destrezas y conceptos amplios, los cuales permitan su aplicación. En cualquier caso, el proyecto PISA no está limitado por el denominador común de lo que se ha enseñado específicamente en cada país participante.

CONOCIMIENTOS Y HABILIDADES PARA LA VIDA PRIMEROS RESULTADOS DE PISA 2000

Los primeros resultados de PISA aportan una línea de base acerca de los conocimientos y habilidades de los estudiantes de cada país participante. A continuación se muestran algunos de los resultados más relevantes para la realidad educativa mexicana.

Cómo se miden los conocimientos y aptitudes en PISA

Las habilidades que se miden en PISA se deben considerar como un continuo: no hay puntos de corte naturales para marcar fronteras entre las etapas de ese continuo. Por lo tanto, es necesario mencionar que dividir el continuo en niveles, aunque útil para la comunicación de los niveles de desempeño, es esencialmente arbitrario.

La división de este continuo en etapas o niveles, sirve para comunicar la competencia de los estudiantes en algo más que números. Esto quiere decir: se debe considerar a un estudiante dentro de un nivel, si la expectativa que se tiene es que resuelva las tareas de ese nivel más que fallarlas. Esto implica que el estudiante obtenga al menos la mitad de reactivos correctos de una prueba de ese nivel. De hecho, la esperanza de éxito del estudiante en un nivel particular, en una prueba que contuviera sólo reactivos de ese nivel, sería de un 50 por ciento para el alumno en la parte baja del nivel y superior a 50 por ciento, para estudiantes más arriba en el mismo nivel.

Para facilitar la interpretación de los puntajes asignados a los estudiantes, PISA ha definido cinco niveles de competencia para lectura, mientras que para matemáticas y ciencias sólo se constituyó una escala con tres puntajes. Dichos puntajes se expresan en una escala que va de 200 a 800 puntos, con un promedio de 500 puntos.

Significado y resultados en los niveles de competencia de lectura

La aptitud para la lectura, se enfoca en el conocimiento y capacidad que se requiere para aplicar el concepto de *leer para aprender* en lugar de tan sólo las destrezas técnicas requeridas para *aprender a leer*.

En línea con los puntos de vista más modernos acerca de la aptitud para la lectura, PISA se concentra en la medición del grado en el cual los individuos son capaces de construir, ampliar y reflexionar, sobre el significado de lo que han leído en una amplia gama de textos comunes, tanto dentro como fuera de la escuela. Los reactivos de lectura más sencillos que pueden, sin embargo, estar asociados con esta noción de aptitud para lectura, son los que se aplican. En el caso de las habilidades para la lectura la descripción de los niveles es como sigue:

Niveles de dominio y su puntaje asociado en la escala de lectura

<u>Nivel</u>	<u>Descripción de competencias</u>	<u>Puntaje</u>
5	En este nivel los estudiantes son capaces de completar reactivos de lectura sofisticados, tales como los relacionados con el manejo de información difícil de encontrar en textos con los que no están familiarizados; mostrar una comprensión detallada de dichos textos e inferir qué información del texto es relevante para el reactivo; así como evaluar críticamente y establecer hipótesis.	Más de 625
4	Los estudiantes son capaces de responder reactivos de lectura difíciles, tales como: ubicar información anidada, interpretar significados a partir de sutilezas del lenguaje y evaluar críticamente un texto.	553 a 625
3	Los estudiantes que dominan el nivel 3 de la escala, son capaces de manipular reactivos de lectura de complejidad moderada, tales como ubicar fragmentos múltiples de información, vincular distintas partes de un texto y relacionarlo con conocimientos familiares cotidianos.	481 a 552
2	Los estudiantes que dominan el nivel 2 de la escala, son capaces de responder reactivos básicos de lectura, tales como ubicar información directa, realizar inferencias sencillas de distintos tipos, determinar lo que significa una parte bien definida de un texto, y emplear cierto nivel de conocimientos externos para comprenderla.	408 a 480
1	Los estudiantes que dominan este nivel son capaces de realizar sólo los reactivos de lectura menos complejos que se han desarrollado para PISA, como ubicar un fragmento de información, identificar el tema principal de un texto o establecer una conexión sencilla con el conocimiento cotidiano.	335 a 407

Dominio en el nivel 5

Solamente el 10 por ciento de estudiantes de todos los países participantes, se mostraron competentes en este nivel. Hay países con más del 15 por ciento de estudiantes en este nivel como Nueva Zelanda, Finlandia y Australia, mientras habría otros con menos del 5 por ciento de su población. México y Brasil, los dos únicos países latinoamericanos que participaron en el estudio del 2000, tienen el 1 por ciento de su población en este nivel.

Dominio en los niveles 4, 3 y 2

En estos tres niveles se encuentran poco menos de las tres cuartas partes de los estudiantes. México tiene un 30 por ciento de su población en el nivel 2, un 19 por ciento en el nivel 3 y otro 6 por ciento en el nivel 4. Sin embargo, debe considerarse que México tiene en el nivel 1 o inferior al 44 por ciento de los estudiantes de su muestra.

Con excepción de Luxemburgo y México, uno de cada cinco estudiantes de los países miembros de la OCDE alcanza el nivel 4 o mayor.

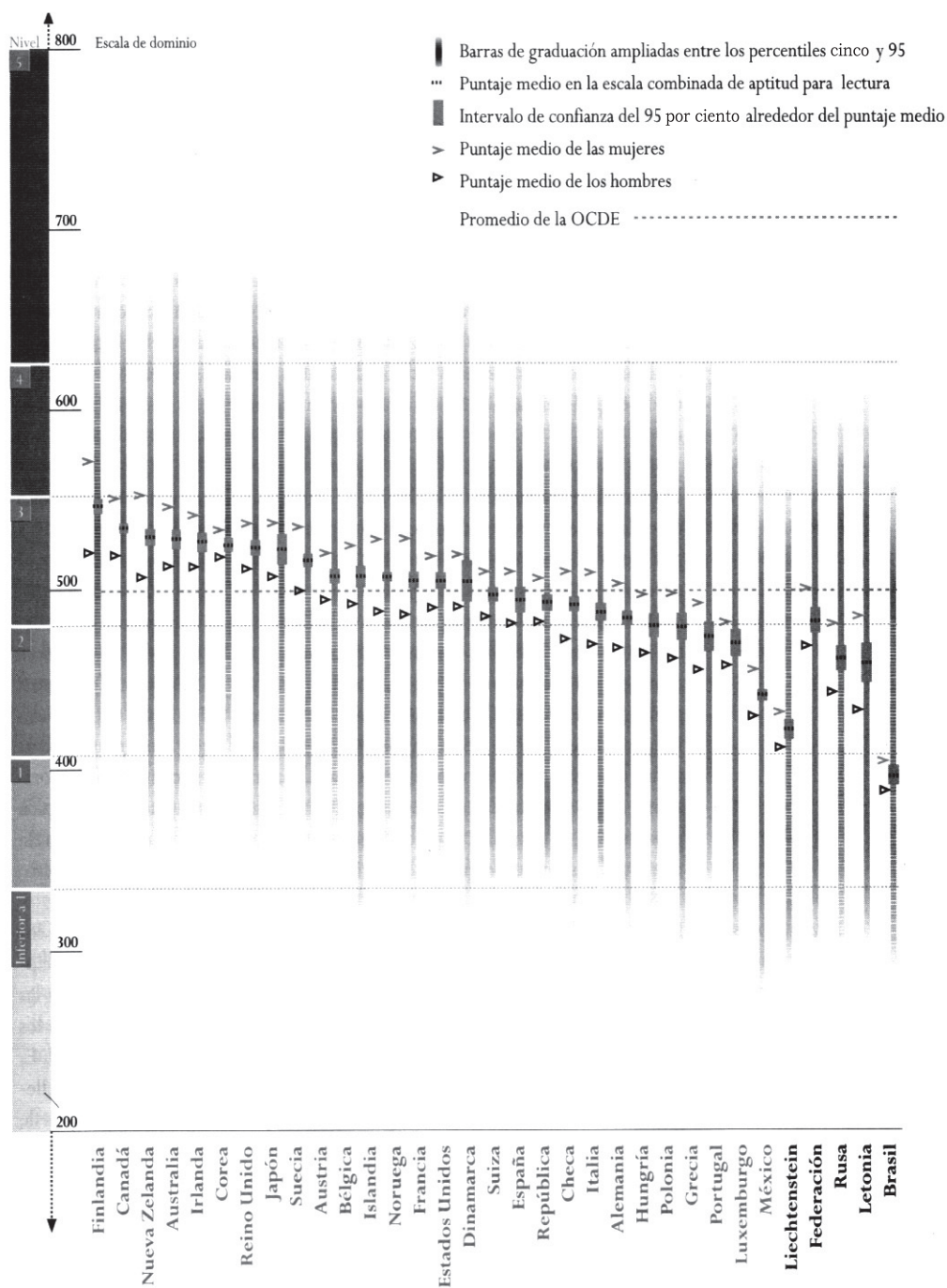
En el área de la OCDE, el 60 por ciento de los estudiantes dominan el nivel 3 (niveles 3, 4 ó 5). En nueve de 27 países el 80 por ciento de los estudiantes de 15 años domina el nivel 3.

En todos los países de la OCDE, la mitad del total de estudiantes se ubica en el nivel 2 o mayor.

Dominio en el nivel 1

En el nivel 1 se encuentra el 12 por ciento de estudiantes evaluados, y un 6 por ciento más no mostró la competencia necesaria para el nivel, situándose por abajo del mismo. Existen marcadas diferencias entre países, habiendo algunos –como Luxemburgo, México y Brasil– donde más del 20 por ciento se ubica en el nivel I o por abajo, mientras países ubicados en los primeros lugares tienen menos del 5 por ciento de sus estudiantes en el mismo nivel.

GRÁFICA 1
La distribución del desempeño de los estudiantes en la escala combinada de aptitud para lectura



Fuente: Base de datos OCDE PISA, 2001. Cuadro 2.3a.

Significado y resultados en los niveles de competencia en matemáticas y ciencias

Dado que la evaluación de las aptitudes para matemáticas y ciencias fue más limitada que la evaluación de la aptitud para lectura en PISA 2000, no se intentó definir niveles de dominio como se hizo para lectura. No obstante, fue posible proporcionar una descripción amplia del desempeño en matemáticas y ciencias, en términos del conocimiento y las habilidades que necesitan demostrar los estudiantes en diversos puntos de las escalas.

Puntajes en las escalas de matemáticas

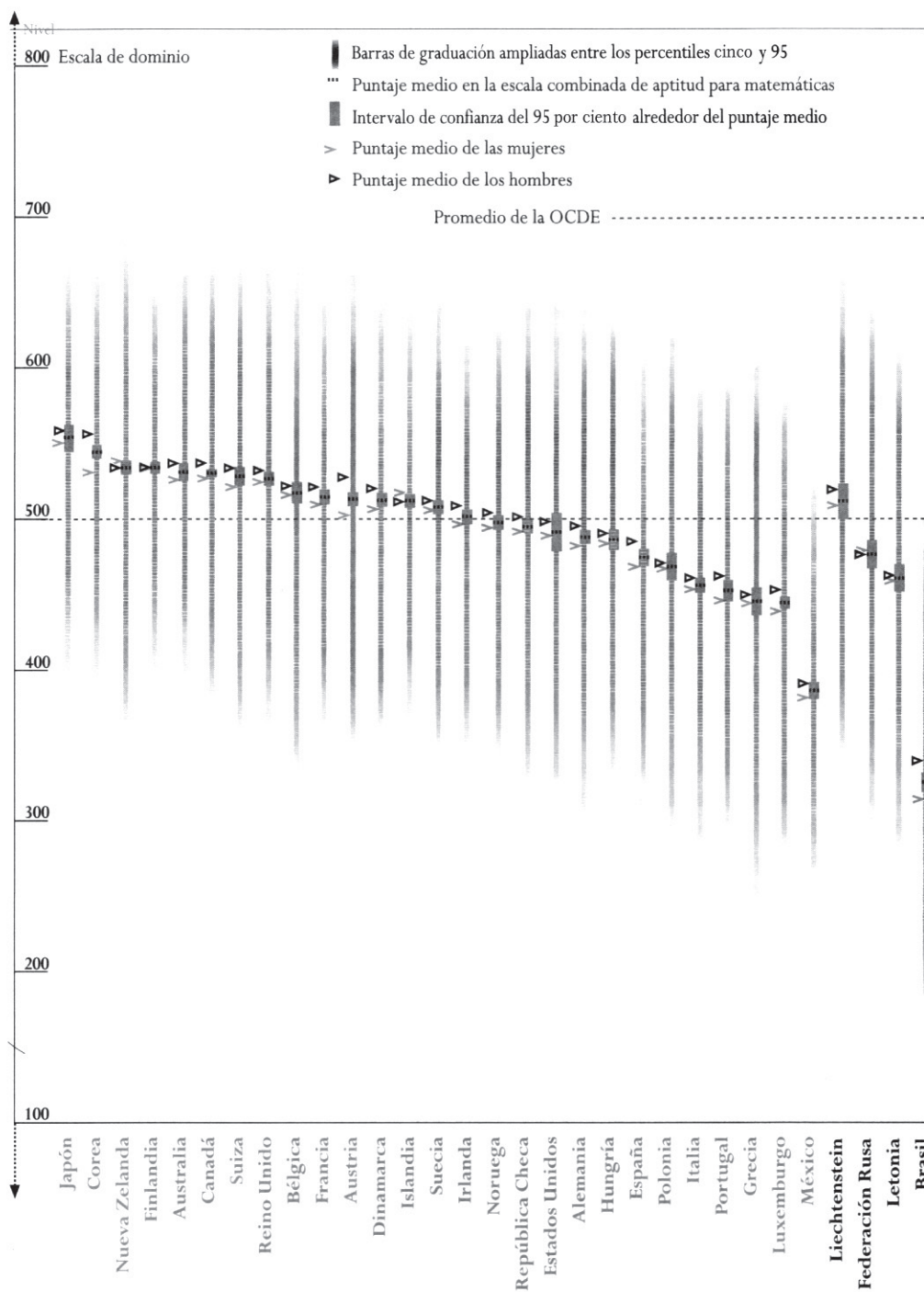
<u>Puntaje</u>	<u>Descripción de competencias</u>
750	Hacia el nivel más alto, los estudiantes normalmente toman una postura activa y creativa en su acercamiento a los problemas matemáticos. Así, interpretan y formulan problemas en términos de matemáticas, son capaces de manejar la información más compleja y de negociar una serie de pasos de procesamiento. Muestran otros procesos cognoscitivos de alto nivel tales como generalización, razonamiento y argumentación para explicar y comunicar resultados.
570	En este nivel de la escala, los estudiantes suelen ser capaces de interpretar, vincular e integrar representaciones distintas de un problema o diferentes fragmentos de información; manipular y emplear un modelo dado, que a menudo involucra el uso de álgebra u otras representaciones simbólicas; así como trabajar con estrategias, modelos o proposiciones dadas.
380	En el extremo bajo de la escala, los estudiantes son generalmente capaces de completar solamente un paso de procesamiento, el cual consiste en la reproducción de elementos matemáticos básicos o en aplicar habilidades simples de cálculo. Cualquier interpretación o razonamiento, casi siempre involucra el reconocimiento de un elemento familiar en el problema. La solución requiere de la aplicación de procedimientos rutinarios en un solo paso de procesamiento.

En la evaluación de PISA, el 5 por ciento más apto de los estudiantes alcanzó en promedio 655 puntos en los países de la OCDE, el 10 por ciento llegó a 625 puntos y el 25 por ciento a 571. Por otra parte, en el extremo bajo de la escala, más de tres cuartas partes alcanzaron por lo menos 435 puntos, más del 90 por ciento llegaron a 367 y más del 95 por ciento tuvieron 326 puntos.

PISA 2000 muestra los desempeños de los estudiantes en matemáticas y en ciencias, sin el nivel de profundidad que se exploró para lectura.

Los resultados en matemáticas permiten observar que Japón logró los resultados promedio más altos (560 puntos), sin separarse notoriamente de Corea y Nueva Zelanda. Por arriba del promedio obtenido por OCDE (500 puntos), también estuvieron Australia, Austria, Bélgica, Canadá y Dinamarca, junto con otros siete países. En el otro extremo se encuentran Brasil y México, con 330 y 380 puntos respectivamente.

GRÁFICA 2
La distribución del desempeño de los estudiantes
en la escala de aptitud para matemáticas



Fuente: Base de datos OCDE PISA, 2001. Cuadro 3.1.

Puntajes en la escala de ciencias

Puntaje	Descripción de competencias
690	Los estudiantes son capaces de crear o emplear modelos conceptuales para hacer predicciones o dar explicaciones; son capaces de analizar investigaciones científicas con objeto de comprender, por ejemplo, el diseño de un experimento o de identificar una idea que se está poniendo a prueba. Comparan datos con el fin de evaluar puntos de vista alternativos o perspectivas distintas, y comunican argumentos científicos o descripciones detalladas y precisas.
550	Los alumnos son normalmente capaces de emplear los conceptos científicos para hacer predicciones o dar explicaciones; de reconocer preguntas que pueden ser respondidas mediante la investigación científica o identificar detalles de lo que está involucrado en una investigación científica y; seleccionar información relevante entre datos que compiten o cadenas de razonamiento para obtener o evaluar conclusiones.
400	En el extremo bajo de la escala, los estudiantes son capaces de recordar conocimientos científicos simples (como nombres, datos, terminología, reglas sencillas) y; emplear el conocimiento científico común para obtener o evaluar conclusiones.

Para ciencias los resultados muestran a Japón y a Corea con los resultados más altos en formación científica, quienes acompañados por Australia, Austria, Canadá, Finlandia, Irlanda, Nueva Zelanda, Reino Unido, República Checa y Suecia, obtuvieron también puntajes significativamente superiores, en términos estadísticos, al promedio de la OCDE para esta área de competencias. México alcanzó 420 puntos y Brasil 370 puntos, ubicándose en los lugares 31 y 32 respectivamente.

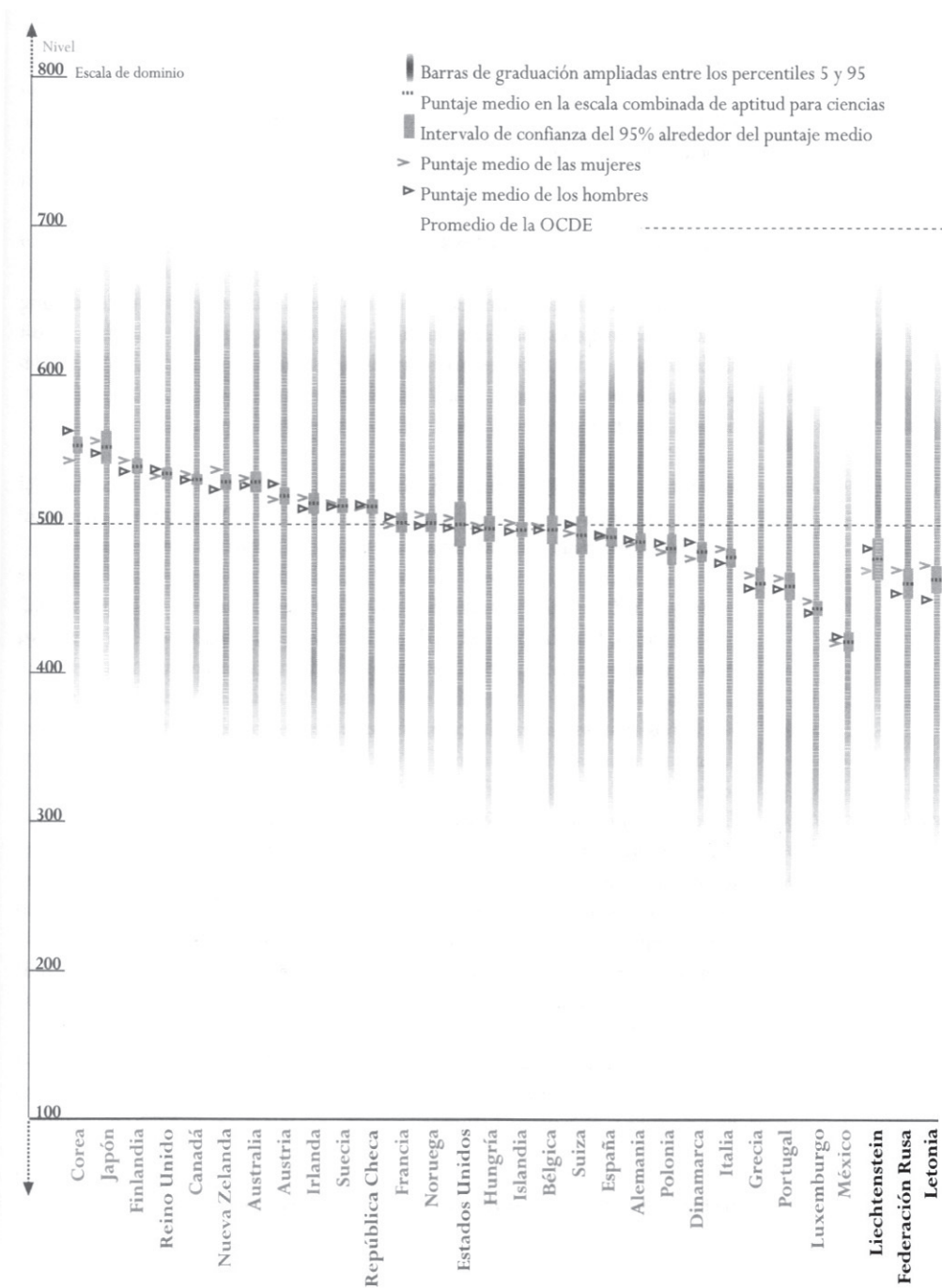
En la escala con puntaje de 500, utilizada también para ciencias, dos terceras partes de los estudiantes en los países de la OCDE alcanzaron entre 400 y 600 puntos.

El 5 por ciento de los estudiantes alcanzó en promedio 657 puntos. El 10 por ciento llegó a 627, y el 25 por ciento alcanzó los 572 puntos.

En el extremo bajo de la escala, más de tres cuartas partes alcanzaron por lo menos 431 puntos. Más del 90 por ciento llegaron a 368, y más del 95 por ciento logró 332 puntos.

Como en el caso de la aptitud para lectura y matemáticas, los resultados en ciencias muestran que la variación en el desempeño estudiantil al interior de los países, es mucho mayor que la variación en el desempeño medio entre naciones. El grado de variación entre éstas cambia considerablemente y el tamaño de la variación al interior de los países no está relacionado con el nivel de su desempeño medio general.

GRÁFICA 3
La distribución del desempeño de los estudiantes
en la escala combinada de aptitud para ciencias



Sobre la interpretación de los resultados

La interpretación de los resultados de PISA debe tomar en cuenta la multidimensionalidad de las pruebas y el traslape de los resultados en el caso de algunos países; así como la cobertura de la muestra respecto a la población objetivo, y a la influencia diferencial del contexto socioeconómico y la escuela sobre los resultados.

La posición que ocupa un país en el conjunto de naciones participantes varía dependiendo del área (lectura, matemáticas y ciencias) y de los aspectos que evalúa cada área, en el caso de lectura: localización, interpretación y reflexión. México, por ejemplo ocupa el lugar 34 en lectura, 35 en Matemáticas y nuevamente el 34 en Ciencias. Considerando los aspectos de lectura, ocupa el lugar 35 en localización, 34 en interpretación y 31 en reflexión.

En los casos de países donde la distancia que separa los promedios obtenidos es menor al margen de error que se maneja, deben considerarse técnicamente empatados, pues no puede asegurarse que uno se encuentre efectivamente arriba o abajo. Tal situación se ilustra bien con el caso de países como Luxemburgo, México y Brasil, cuyos puntajes los colocan en los últimos lugares, y Finlandia ubicado en el primer sitio: habría un 95 por ciento de confianza sobre la corrección de sus respectivos lugares; sin embargo, no es el caso de Canadá, quien estando en el segundo lugar podría también ocupar el tercero o cuarto.

Los aspectos anteriores muestran los límites de las interpretaciones en términos de rankings y la falta de solidez y respaldo inherentes a ciertas aseveraciones que se hacen a partir de las mismas. Sobre todo en el ámbito periodístico, donde los encabezados tienen una máxima economía de palabras, por lo que resulta fácil e impactante decir: México, penúltimo lugar del mundo, sin considerar múltiples elementos que condicionan el planteamiento. Por ejemplo, no todos los países fueron evaluados, no todos presentan el mismo nivel de desarrollo; los sistemas educativos de cada país difieren por cuestiones históricas, culturales, financieras, etcétera.

El Instituto Nacional para la Evaluación de la Educación (INEE), considera importante fijar su posición respecto a la ética profesional que conlleva la evaluación educativa, y recordar algunos aspectos vinculados a la actitud intelectual que debe dirigir la interpretación y comunicación de la información mostrada públicamente, a propósito de los resultados de PISA.

La evaluación compara datos empíricos o evidencias sobre la realidad estudiada, respecto a ciertos referentes que permiten detectar el nivel de alejamiento o acercamiento de los datos encontrados. La calidad y pertinencia de cualquier estudio de evaluación están asociadas a un manejo adecuado de los datos empíricos, respetando las normas que la comunidad científica ha elaborado para tal efecto, pero también de la selección de los referentes utilizados para la comparación de la información.

En este punto se toca un aspecto primordial que corresponde a la razón de ser de la evaluación y le confiere identidad: contribuir al mejoramiento de las prácticas educativas y logro de los estudiantes.

METODOLOGÍA PARA EL DISEÑO Y APLICACIÓN DE LAS PRUEBAS PISA

La medición por medio de pruebas estandarizadas a gran escala en diferentes países, como el estudio de PISA, comprende una metodología altamente compleja y con una amplia e intensa colaboración entre diferentes niveles de responsabilidad, tanto nacionales como internacionales.

Debido al alcance del estudio y a la necesidad de obtener datos de alta calidad para poder apoyar los objetivos del estudio, la metodología de PISA se caracteriza por ser integral, sistemática y rigurosa. Constituye una guía de actuación al definir las normas de diseño, procedimiento y análisis del ciclo del estudio.

DISEÑO DE INSTRUMENTOS

Diseño de las pruebas

- *Definición del marco de referencia de los dominios de evaluación*

La función del marco de referencia es dirigir las evaluaciones de PISA. En éste se define cada dominio, describe el alcance de la evaluación, el número de reactivos que se necesitan para medir cada componente del dominio, determina el balance de los tipos de preguntas y bosqueja el tipo de reporte de resultados.

- *Desarrollo de pruebas*

El grupo responsable para el desarrollo de las pruebas se integra por personal del *Australian Council for Educational Research (ACER)* y de *CITI Group*, además participan miembros de los FEGs, traductores y revisores de reactivos de los centros nacionales, quienes contribuyen sustancialmente al proceso para el desarrollo de las pruebas. Abarca las siguientes actividades:

- Convocatoria para el diseño de reactivos.* El consorcio convoca a los países participantes a enviar reactivos, de acuerdo con las guías elaboradas *ex profeso*.
- Diseño de reactivos y revisión técnica y de congruencia.* El grupo encargado del diseño de las pruebas revisa los reactivos enviados por los países y se asegura que sean consistentes con el marco de referencia, y que técnicamente estén bien diseñados.
- Revisión de reactivos por un grupo de expertos.* El proceso de revisión se realiza, primero por parte de expertos, después en los países participantes, y finalmente, por los FEGs de cada dominio.

- d) *Preparación de las guías de calificación y material de entrenamiento para los calificadores.* Se preparan dos tipos de materiales para el calificador: la Guía de Calificación y el Manual de Capacitación de cada dominio. La guía incluye la forma de asignar los códigos a las respuestas, y el manual sirve para ejercitar a los calificadores en su tarea de asignación de códigos.
- f) *Prueba piloto, selección, revisión técnica y de expertos de los reactivos para el estudio; y revisión cultural.* En promedio, cada reactivo es respondido por 150 ó 200 estudiantes por país en la prueba piloto. La revisión y selección de los reactivos para la prueba definitiva es realizada por los revisores nacionales, el grupo de diseño de la prueba, los FEGs, y el grupo de expertos encargado de la revisión cultural de los reactivos.
- g) *La conformación del banco de reactivos e integración de pruebas.* El banco de reactivos de las pruebas almacena y permite la selección de los reactivos. Después de la prueba piloto, se obtiene y captura la información estadística de los reactivos, como los índices de dificultad y de discriminación.
- *Diseño de los cuestionarios de contexto*
Para el estudio de PISA se emplean instrumentos de contexto, tales como el Cuestionario del Estudiante y el Cuestionario para la Escuela. Su función es obtener datos que sirvan para construir indicadores de factores social, cultural, económico y educativo, que se presume tengan influencia o estén asociados al desempeño de los estudiantes. El diseño de estos instrumentos implica un proceso similar al seguido para las pruebas:

ASPECTOS OPERATIVOS

Diseño de la muestra

La población objetivo para el estudio de PISA comprende estudiantes de 15 años inscritos de tiempo completo o parcial en las escuelas de cada país. Los estándares de calidad que PISA marca para el muestreo se refieren a: cobertura de la población objetivo, precisión y confiabilidad; y tasas mínimas de respuesta de escuelas y estudiantes.

Traducción y adecuación de las pruebas y del material del estudio

La causa principal de un comportamiento pobre de los reactivos en las pruebas internacionales son los errores de traducción. Por ello, PISA desarrolló procedimientos estrictos de traducción y equivalencia, los cuales son evaluados antes de aplicar la prueba.

Doble traducción proveniente de dos idiomas fuente

La verificación de las equivalencias entre los idiomas *fuentes* (inglés y francés) y la traducción para la población objetivo, es realizada por cuatro diferentes personas, dos traductores, un conciliador nacional y un verificador del consorcio. Uno de los controles de calidad más productivos implantados por PISA para mantener los altos estándares en la traducción de los materiales de evaluación consiste en la formación de un equipo de traductores independientes nombrados y entrenados por el consorcio, para revisar cada versión nacional contra las versiones en inglés y francés.

OPERACIONES DE CAMPO

Capacitación internacional

Representantes de cada país requieren ser sometidos a dos sesiones de capacitación, una inmediatamente antes del proceso de campo, y la otra inmediatamente antes del estudio principal. Durante estas sesiones de entrenamiento, el consorcio familiariza al equipo nacional con las interpretaciones de las Guías de Calificación y los Manuales de Capacitación.

Equipo de codificadores

El NPM es el responsable de contratar el personal encargado para la codificación individual y múltiple. El número de codificadores requeridos se define por el modelo del procedimiento de la codificación y por el tamaño de la muestra seleccionada.

Cómo se hace la codificación

Una línea de números aparece en la parte superior derecha de cada reactivo en los cuadernillos; corresponde a los posibles códigos, los cuales se pueden aplicar a cada respuesta del estudiante en los casos de reactivos de respuesta abierta. La codificación es realizada directamente en el cuadernillo, circulando el número seleccionado del código, de conformidad con lo dispuesto en la Guía de Calificación para la respuesta de cada estudiante en específico.

Captura de datos, revisión y entrega de archivos

El Consorcio entrega al país participante un paquete para captura de datos (*Key Quest*), el cual contiene la base de datos estructurada para registrar la información de los cuader-

nillos y los cuestionarios. El NPM es el responsable de asegurar la calidad de los datos del país, antes de que los archivos sean entregados al consorcio.

Monitoreo de la calidad

Es esencial que un proyecto con un perfil tan alto y costoso como el proyecto PISA, deba realizarse con altos estándares de calidad. PISA prepara monitores de calidad, los cuales observan las operaciones desde el proceso de campo, sin embargo, la responsabilidad del control de calidad recae en el NPM, quien instrumenta y guía la operación.

PROCESAMIENTO DE DATOS

Ponderación del estudio y cálculo de la varianza de la muestra

Por el tipo de estudio que representa PISA, el procesamiento de datos se constituye en una fase primordial para generar información confiable, válida y significativa para los países participantes.

Un tipo de procesamiento es la ponderación del estudio, el cual permite calcular el error de muestreo y determinar la validez de las estimaciones e inferencias. Este procesamiento se dirige al análisis de las muestras para mejorar las estimaciones y lograr una mayor confiabilidad en las interpretaciones. La ponderación es necesaria, debido a que los estudiantes incluidos en las muestras del estudio de cada país, no son igualmente representativas de la población total, a pesar de haber realizado el muestreo al azar de escuelas y estudiantes.

Calibración de los reactivos

Otro análisis realizado se centra en la validez y confiabilidad de los instrumentos de medición, para lo cual se efectúa un análisis de reactivos bajo el modelo de *Respuesta al Item de Rash*. Este modelo se aplica para la calibración nacional e internacional, así como para la obtención de los puntajes de los estudiantes.

CONSTRUCCIÓN DE ESCALAS

Construcción de escalas de competencias

PISA pretende reportar resultados de competencias que estén basadas en una teoría científica, y que dichos resultados sean interpretables en términos de políticas educativas. Para ello desarrolla escalas que reflejan, por niveles, los resultados de los estudiantes eva-

luados. PISA debe entregar resultados individuales para cada país y por cada uno de los dominios cognoscitivos evaluados, así como entregar la descripción de las competencias, la cual debe arrojar luz sobre las tendencias a través del tiempo.

Desarrollo de la descripción de las escalas

El desarrollo de la descripción de las escalas de competencias PISA, es un proceso compuesto por siete fases: identificación de posibles subescalas, asignación de reactivos a las escalas; auditoría de habilidades; análisis de los datos del proceso de campo; definición de las dimensiones; refinación de los datos del estudio principal, y validación.

Definición de los niveles de competencia

La definición de los niveles de competencia implica dos etapas: el análisis sustantivo de los reactivos PISA, en relación con los aspectos cognoscitivos que soporta cada dominio. Esto produce niveles de competencia, los cuales reflejan las observaciones del desempeño de estudiantes, y un análisis detallado de las exigencias cognoscitivas de los reactivos de PISA. La segunda, comprende decisiones en torno a la ubicación de los puntos de corte para los niveles y la forma de asociar estudiantes a cada nivel. Esta etapa permite interpretar lo que significa *estar en un nivel* y tiene consecuencias sustanciales en los reportes de resultados nacionales e internacionales.

PISA 2003 EN MÉXICO

Desde la primera evaluación de PISA, México ha estado involucrado en la implementación de las pruebas piloto y las aplicaciones. Hasta ahora ha sido uno más de los países que se suman al proyecto PISA, con el convencimiento de que someterse a mediciones de esta envergadura, es crucial para la identificación comparativa de áreas débiles y fuertes en la educación básica, y sobre todo, para la promoción y establecimiento de nuevas y mejores políticas educativas.

APLICACIÓN

Capacitación

El 10 y 11 de abril de 2003, se realizó una reunión para capacitar a los Coordinadores Estatales de la SEP, sobre la aplicación del proyecto PISA 2003. El contenido de la capacitación abarcó las funciones de los aplicadores; actividades a realizar antes, durante y después de

la aplicación; instrucciones para el llenado de las Cédulas de Registro de Alumnos; así como la programación concerniente a:

- Periodo de la aplicación,
- Entrega de las listas de alumnos seleccionados conforme a la muestra de escuelas,
- Entrega del calendario de la aplicación,
- Envío a las representaciones de los recursos financieros y de los materiales de la aplicación.

Preparación de la aplicación

Después de la capacitación se llevaron a cabo diversas tareas, las principales fueron:

- De abril a mayo se tradujeron al español los cuadernillos de conocimientos. Igualmente, se revisaron y ajustaron los cuestionarios de contexto, y los manuales del aplicador y del coordinador escolar, que ya habían sido traducidos al español.
- Entre abril y mayo se empezaron a recibir las bases de datos con la información de las listas de los alumnos de 15 años, por parte de las representaciones estatales. Se fueron corrigiendo y depurando las bases, hasta tenerlas limpias para generar las Cédulas de Registro del Alumno.
- La entrega de los materiales de aplicación a las representaciones se realizó a partir del 21 de mayo. La población a quien se le envió material de evaluación constaba de 35 mil 730 alumnos de mil 170 escuelas.

Población evaluada (alumnos y escuelas)

Las pruebas PISA se aplicaron a un total de 30 mil 087 alumnos de 15 años de edad, de mil 131 escuelas, de 31 de las 32 entidades federativas, durante los días 29 y 30 de mayo, y del 2 al 6 de junio del 2003. No se llevó a cabo la aplicación en Michoacán, debido a problemas de paro en las escuelas del estado.

Al revisar la distribución de población evaluada, destaca que el mínimo de escuelas y de alumnos evaluados se registró en Campeche. Por otro lado, el máximo de escuelas evaluadas se llevó a cabo en Veracruz; y el máximo de alumnos, en Yucatán.

La tasa de participación de alumnos fue de 84 por ciento, en tanto que la de escuelas fue de 97 por ciento (tasa con reemplazo). De acuerdo con el *School Sampling Preparation Manual* las tasas mínimas de participación requeridas por PISA para las escuelas es de 85 por ciento y para alumnos de 80 por ciento. La comparación entre las tasas de participación obtenidas y requeridas se muestra en el siguiente cuadro.

Tasas de participación obtenidas y requeridas

	Tasa de participación	
	Obtenida	Requerida
Alumnos	84%	80%
Escuelas	97%	85%

Si se analizan las tasas de participación por entidad, se detecta que en Campeche, Nuevo León, y Oaxaca las tasas de participación de escuelas fueron inferiores al 85 por ciento. Las entidades con una tasa de participación de alumnos inferior al 80 por ciento fueron Campeche, Distrito Federal, Nuevo León, Oaxaca, Sinaloa y Veracruz.

Recepción de material

Durante la semana del 9 al 13 de junio del 2003, las representaciones estatales de la SEP enviaron cajas con materiales de evaluación. Las cajas fueron almacenadas en la bodega del INEE, en donde se contaron los cuadernillos de conocimientos, los cuestionarios para el alumno y la escuela, y se separó el material utilizado del no utilizado.

PROCESO DE CODIFICACIÓN

Traducción y preparación de guías y manuales

A partir de la primera quincena de junio y hasta julio de 2003, se tradujeron las cuatro Guías de Calificación y los cuatro Manuales de Capacitación, correspondientes a matemáticas, solución de problemas, ciencias y lectura. Las traducciones fueron enviadas a PISA para su revisión y aprobación en la última semana de agosto. Las observaciones remitidas por PISA se recibieron la siguiente semana y fueron incorporadas a las guías y los manuales.

Selección y contratación de codificadores y supervisores

Antes de la contratación, se determinó el ordenamiento de lectura y matemáticas, y conforme a éste se tomaron las siguientes decisiones:

- Disponer de dos grupos de codificadores, uno para lectura y el otro para matemáticas, solución de problemas y ciencias,
- Contratar para lectura veinte personas: 16 calificadores y cuatro supervisores;

y para matemáticas 45 personas: cuarenta calificadores y cinco supervisores,

- Integrar para lectura cuatro grupos de cuatro personas cada uno y para matemáticas diez grupos de cuatro personas,
- Seleccionar personas de reserva para que en caso de que alguno de los contratados no asistiera, se pudiera sustituir,
- El proceso de codificación sería desfasado, al iniciar con lectura que empezaría el 5 de septiembre y después matemáticas, que iniciaría el 17 de septiembre,
- Citar a las supervisoras un día antes del inicio de la calificación, a fin de capacitarlas en el área respectiva y explicarles sus funciones.

Durante la primera quincena de agosto se definió el perfil del codificador. Se usó el recurso de la Bolsa Universitaria de Trabajo de la UNAM para solicitar personal conforme al perfil definido. La oferta de trabajo se publicó en internet el 11 de agosto y estuvo disponible durante 15 días. A partir del 12 de agosto se inició el proceso de recepción de solicitudes y la concertación de citas para entrevistas.

Las entrevistas se realizaron del 19 al 28 de agosto, en donde se seleccionó a los codificadores y se definió la contratación. En total se contrataron 15 personas para el área de lectura, 12 como calificadores y tres como supervisoras. Para el área de matemáticas, solución de problemas y ciencias se contrataron 39 personas, 36 como calificadores y tres para fungir como supervisoras.

Selección de los cuadernillos para la codificación múltiple

Con la base de datos ligada de los materiales enviados con las Cédulas de Identificación de los Estudiantes, se seleccionaron los 900 cuadernillos de la codificación múltiple, cien de cada versión (cuadernillos del 1 al 6, 8, 10 y 12). Una vez seleccionados electrónicamente, se extrajeron de las cajas respectivas para ser colocados aparte.

CODIFICACIÓN

Lectura

La codificación se inició con lectura el 5 de septiembre del 2003. Ese día se llevó a cabo la capacitación de los 12 codificadores, por parte del NPM de México. Los codificadores contaban con su Guía de Calificación y el Manual de Capacitación. La jornada era de 9 a 15:30 hrs. con un receso de 15 minutos a la mitad de la misma.

Se integraron tres grupos de cuatro personas cada uno y las dos supervisoras atendían a dos grupos por cada una de ellas.

La distribución de las versiones de los cuadernillos por codificador atendió las reglas de ordenamiento dadas en el capítulo 7 del *National Program Manager Manual*. Para el ordenamiento de México intervino la variable entidad para la distribución de los cuadernillos. Se distribuyeron los integrantes de los tres grupos por entidad y versiones de cuadernillo, tanto para Lectura 1 como para Lectura 2.

La dinámica diaria de codificación consistía en la llegada de los codificadores y supervisoras, entrega de cajas con los cuadernillos a los codificadores, codificar en su mesa y grupo respectivo, atención de dudas y revisión de cuadernillos por parte de las supervisoras; conforme se iba terminando de codificar cada caja se regresaba al almacén y se entregaba una nueva. Antes de finalizar la jornada diaria, personal del *staff* encargado de la operación, revisaba una muestra de cuadernillos para verificar la calidad de la codificación y se reunía con las supervisoras para analizar la jornada y, si era el caso, dar indicaciones para el siguiente día.

El material de apoyo que estuvo disponible para los codificadores fue la Guía de Calificación, y para las supervisoras fue la distribución de versiones de cuadernillos y entidades por calificador; una lista de los reactivos por módulo, el rango de páginas de los módulos por cuadernillo y la Guía de Calificación.

Para la calificación múltiple, se seleccionaron siete calificadores, tres para la codificación individual (simple marking) de la múltiple; y cuatro para la codificación múltiple, este último proceso se efectuó el 25 y 26 de septiembre.

Matemáticas, solución de problemas y ciencias

La codificación se inició el 17 de septiembre, y ese día se capacitó en los dos primeros módulos de matemáticas. Las capacitaciones siguientes fueron el 22 (módulos 3-7 de matemáticas) y 30 de septiembre (módulos 1 y 2 de solución de problemas), y el 3 de octubre (módulos 1 y 2 de ciencias). La jornada fue igual que para lectura. Se integraron nueve grupos de cuatro personas cada uno y las tres supervisoras atendían a tres grupos por cada una de ellas.

Como en lectura, la distribución de los cuadernillos por codificador siguió las reglas de ordenamiento del capítulo 7 del *National Program Manager Manual*. Para el ordenamiento de matemáticas, solución de problemas y ciencias, intervino la variable entidad para la distribución de los cuadernillos. Se distribuyeron los integrantes de los nueve grupos por entidad y versiones de cuadernillo para los siete módulos de matemáticas, los dos de solución de problemas y los dos de ciencias. En esta codificación se asumió como un dominio integral a las tres áreas.

La dinámica y el material de apoyo fue similar que en lectura. Para la calificación múltiple, se seleccionaron 16 codificadores. El proceso de codificación múltiple se efectuó del 13 al 14 de octubre.

Digitación

A finales de septiembre se inició la digitación de los cuestionarios de contexto aplicados. Se digitaron alrededor de 100 mil documentos entre los cuestionarios, cédulas de registro de alumnos, cuadernillos, reporte de aplicación y lista de programas de estudio. Esto implicó más de veinte millones de golpes de máquina a digitar.

CONCLUSIONES

En los últimos años, México ha centrado sus políticas públicas de carácter social—particularmente las educativas— en enfrentar problemas acumulados de rezago y marginación, con la convicción de que el desarrollo del país pasa por el bienestar de sus habitantes, y hablar de bienestar es plantearse la cuestión de la justicia social, equidad y calidad de las oportunidades que constitucionalmente se encuentran garantizadas para todos los mexicanos.

Actualmente están siendo operadas solamente para el subprograma sectorial de educación básica, 16 políticas educativas que concentran 41 programas o proyectos precisamente dirigidos a la equidad y calidad del proceso y el logro educativos, así como a la reforma de la gestión institucional, como plataforma de apoyo para mejorar el desempeño integral del sistema.

El comportamiento de los diferentes países en los resultados de PISA, muestra la posibilidad de intervención que tienen los gobiernos, para que a través de políticas públicas dirigidas a lograr la efectividad de las escuelas, éstas contribuyan a mejorar el rendimiento escolar. En este marco, las acciones educativas toman en cuenta las desventajas con las cuales acceden los estudiantes a las escuelas, homogeneizando los resultados hacia niveles más altos de efectividad.

La primera lección que ofrece PISA, es la confirmación de que los esfuerzos de México por atender la equidad y calidad de su sistema educativo deben ser secundados y reforzados con estrategias de mayor impacto. Si bien los cambios que involucran factores cualitativos demandan madurez y cuidado especial en los procesos, y por tanto tiempo y planeación rigurosa, hay aspectos que no puedan esperar, los cuales requieren atención urgente. La evaluación es uno de estos aspectos, por la importancia que reviste como generadora de conocimiento sobre el sistema educativo.

De ello se deriva la necesidad de intensificar el conocimiento de la realidad del país, cuestionando las certezas e integrando diferentes perspectivas, de manera que se logren acercamientos rigurosos sobre los fenómenos que nos preocupan y se elijan las herramientas adecuadas para enfrentarlos. En este sentido, la evaluación ofrece información útil para explicar la cuestión educativa, razón por la cual se requiere:

- ▲ Construir y aplicar instrumentos de probada calidad y equiparables a los instrumentos internacionales, de manera que los resultados ofrezcan la posibilidad de la comparación.
- ▲ Encontrar caminos que permitan ampliar las muestras de estudiantes y desarrollar análisis con diferentes niveles de desagregación.
- ▲ Incrementar los estudios de carácter cualitativo en combinación con los estudios basados en instrumentos objetivos que favorecen las aplicaciones de gran escala.
- ▲ Intensificar la participación en evaluaciones internacionales, las cuales permitan validar y mejorar las evaluaciones nacionales, así como contar con referentes externos que aporten conocimientos para la mejora interna.
- ▲ Incrementar esfuerzos para que los resultados de las evaluaciones lleguen a los diferentes actores educativos y obtengan provecho de los mismos.

México ha participado en las aplicaciones y pruebas piloto realizadas del 2000 a la fecha, y ha cumplido a cabalidad con los estándares de aplicación y muestreo requeridos por PISA para realizar un estudio comparativo de índole internacional.

Una de las áreas en donde México debe intervenir activamente, es el diseño de reactivos de acuerdo con el marco definido por PISA. Con la experiencia en evaluación y la participación en este estudio, México, a través del INEE, puede promover la investigación en las competencias básicas de razonamiento verbal y matemático en la población de educación básica, la evaluación estandarizada; y conformar grupos de trabajo dedicados al diseño de reactivos, que después de ser probados en México y Latinoamérica, sean incluidos en las pruebas de PISA.

Es prioritario para México continuar participando en este tipo de estudios comparativos a nivel internacional, dada la importancia que revisten los resultados y el impacto que pueden tener en las políticas de largo aliento implantadas en la educación básica.

Además, los resultados servirán como fuente de investigación nacional, y permitirán el desarrollo de estudios específicos que ayuden al entendimiento y mejoramiento de la educación. Estos estudios propiciarán la detección de situaciones problema y se podrán proponer proyectos estratégicos que contribuyan a la mejora continua de la educación nacional.