



Instituto Nacional para la  
Evaluación de la Educación


---

# **El Proyecto PISA: Su Aplicación en México**

---

Mtro. Rafael Vidal Uribe  
Lic. Antonieta Díaz González  
Lic. Javier Noyola  
*Dirección de Proyectos Internacionales y Especiales, del INEE*

---

 CONTENIDO	Pág.
Introducción	3
1. El proyecto PISA	6
2. Conocimientos y Habilidades para la Vida. Primeros resultados de PISA 2000	16
1. Cómo se miden los conocimientos y aptitudes en PISA	17
1.1 Significado de los niveles de competencia en lectura	18
1.2 Significado de los niveles de competencia en matemáticas y ciencias	22
2. Resultados en lectura	23
3. Resultados en matemáticas	26
4. Resultados en ciencias	27
5. Conocimientos y habilidades para la vida en relación con el contexto	28
6. PISA 2000 “Aptitudes básicas para el mundo del mañana”	30
6.1 Lectura	30
6.2 Matemáticas y ciencias	31
7. Sobre la interpretación de los resultados	32
8. Los factores que influyen en la calidad del rendimiento escolar	35
9. La lucha de México por la equidad con la calidad. Lecciones que aporta la evaluación de PISA	38

CONTENIDO	Pág.
3. Metodología para el diseño y aplicación de las pruebas PISA	40
1. Diseño de Instrumentos	42
1.1 Diseño de pruebas	42
1.2 Diseño de los cuestionarios de contexto	54
2. Aspectos Operativos	58
2.1 Diseño de la muestra	58
2.2 Traducción y adecuación de las pruebas y del material de estudio	61
2.3 Operaciones de campo	64
2.4 Monitoreo de calidad	81
3. Procesamiento de datos	85
3.1 Ponderación del estudio y cálculo de la varianza de la muestra	85
3.2 Calibración de los reactivos	87
3.3 Codificación y estudios de confiabilidad de los calificadores	89
3.4 Procedimientos de limpieza de datos	91
4. Construcción de escalas	94
4.1 Construcción de escalas de competencia	93
4.2 Base de datos internacional	103
4. PISA 2003 en México	107
1.1 Aplicación	108
1.2 Proceso de codificación	110
1.3 Digitación	117
Referencias	118
Siglas	120

# Introducción

¿Están 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? Estas son preguntas para las que los padres de familia, los maestros, las autoridades educativas, los estudiantes y el público en general deberían tener una respuesta.

Casi todos los sistemas educativos suelen monitorear en mayor o menor medida el aprendizaje de los alumnos con el fin de obtener una imagen de la situación nacional y con ello impactar a la política educativa nacional. Un estudio comparativo internacional del rendimiento de los estudiantes es una herramienta invaluable para extender y enriquecer el panorama nacional al proveer de un contexto más grande para interpretar los resultados nacionales.

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 (de dentro y fuera de la OCDE) en términos de logros de los estudiantes. Al proporcionar información relevante de los resultados educativos y de los contextos en los que éstos ocurren, PISA se propone fundamentalmente ayudar a reflexionar a los países participantes sobre la política y las metas educativas y 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 es la 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 más a evaluar si es que los estudiantes pueden aplicar el conocimiento que han aprendido en la escuela 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 estudio PISA y sobre la información que debe ser recabada están a cargo de expertos de reconocido prestigio internacional. Existe asimismo 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 etc., son monitoreados regularmente por agentes externos a las instancias responsables de operarlo. Como consecuencia de lo anterior, los resultados de PISA tienen un alto grado de validez y confiabilidad y pueden ayudar a mejorar significativamente nuestra comprensión de los resultados de la educación en los países que participan en el proyecto.

Es sabido que el establecimiento de programas educativos de buena calidad es un aspecto fundamental para dar a los jóvenes un buen impulso para promover en ellos el aprendizaje a lo largo de la vida y que si esto no ocurre en las etapas tempranas de la vida es muy difícil remediarlo después y no sólo se daña a los individuos sino también a la

sociedad; sin embargo, en muchos países el establecimiento de políticas educativas adecuadas y programas de cobertura son esfuerzos aislados o fragmentados; por ello, es muy 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.

El presente informe pretende mostrar el qué y el para qué del proyecto PISA, y difundir algunos de sus resultados más importantes y la manera como este tipo de estudios se desarrollan y lo que se ha hecho en México para llevarlo a cabo. El informe está organizado en cuatro capítulos. El *primero* es una visión general de lo qué es el proyecto PISA, su enfoque y la manera cómo está organizada su operación. El *segundo* es una síntesis de los resultados más importantes que ha arrojado la primera evaluación y sus implicaciones para la realidad mexicana. El *tercero* es una exposición de la manera como el proyecto PISA se ha llevado a cabo. El *cuarto*, son las particularidades y datos sobre la aplicación del proyecto PISA en México.

El INEE presenta en este espacio las características y conclusiones más relevantes del proyecto PISA con la intención de ofrecer elementos para una comprensión integral de sus propósitos y resultados, pues se tiene la firme convicción de que en la medida que se comprenda mejor esto será de más utilidad para ayudar a mejorar la educación mexicana.

# 1

## El proyecto PISA

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

El proyecto parte del supuesto que obtener datos sobre el desempeño de los estudiantes, su contexto individual, la familia y los factores institucionales y sociales que influyen en el desempeño y conocer su dinámica es de gran ayuda para explicar las diferencias en el desempeño de los estudiantes. PISA se enfoca más a evaluar las habilidades para usar los conocimientos y destrezas para enfrentar retos de la vida real que a la evaluación de la medida en que los estudiantes han logrado adquirir un currículo escolar específico.

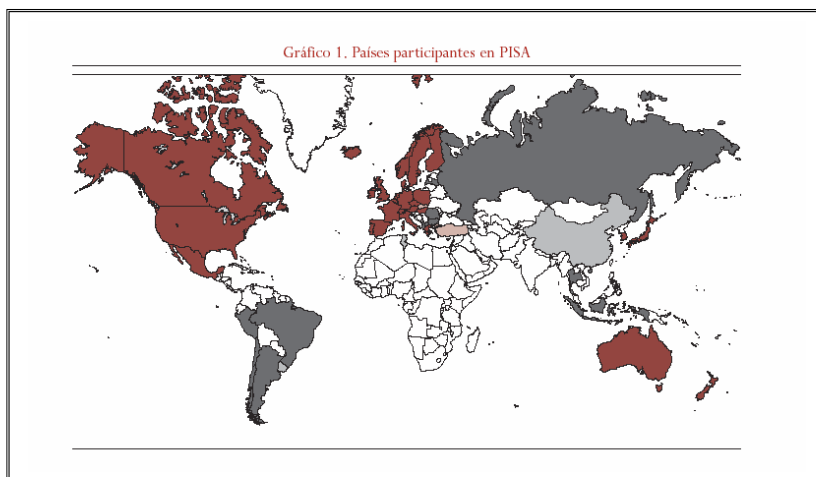
Las principales características del estudio son:

- Su orientación expresa 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.
- Su enfoque innovador hacia la alfabetización no sólo de la lectura sino también hacia las ciencias y las matemáticas.
- Su enfoque en la demostración de conocimientos y habilidades de manera que sea relevante a la vida diaria.
- Su amplitud en la cobertura geográfica con 45 países participando lo que representa un tercio de la población mundial.
- Su regularidad con un compromiso de repetir el estudio cada tres años.
- Su 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 4 países no miembros). En el año 2002 otros 12 países no miembros de la OCDE llevaron a cabo una aplicación complementaria; sin embargo 3 países (Holanda, Rumania y China) no pudieron estar presentes en el estudio con lo que se llegó 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*.

En el año 2003 se ha realizado una segunda aplicación del proyecto PISA en la que han participado 41 países (Ver Tablas).



**PAÍSES PARTICIPANTES EN PISA 2000**

<i>Países de la OCDE participantes en PISA 2000</i>		<i>Países no miembros de la OCDE participantes en PISA 2000</i>	<i>Países donde la evaluación se completó en 2002</i>
Australia	Japón	Brasil	Albania
Austria	Corea	Latvia	Argentina
Bélgica	Luxemburgo	Liechtenstein	Bulgaria
Canadá	México	Federación Rusa	Chile
República Checa	Holanda		China
Dinamarca	Nueva Zelanda		Regional de Hong Kong
Finlandia	Noruega		Indonesia
Francia	Polonia		Israel
Alemania	Portugal		Macedonia
Grecia	España		Perú
Hungría	Suecia		Rumania
Islandia	Suiza		Tailandia
Irlanda	Reino Unido		
Italia	Estados Unidos		

### PAÍSES PARTICIPANTES EN PISA 2003

<i>Países de la OCDE participantes en PISA 2000</i>		<i>Países no miembros de la OCDE participantes en PISA 2003</i>
Australia	Corea	Brasil
Austria	Luxemburgo	Hong Kong China
Bélgica	México	Indonesia
Canadá	Holanda	Latvia
República Checa	Nueva Zelanda	Macau China
Dinamarca	Noruega	Perú
Finlandia	Polonia	Federación Rusa
Francia	Portugal	Serbia
Alemania	República Eslovaca	Tailandia
Grecia	España	Túnez
Hungría	Suecia	Uruguay
Islandia	Suiza	
Irlanda	Turquía	
Italia	Reino Unido	
Japón	Estados Unidos	

### Periodicidad

Está programado que el estudio se lleve a cabo regularmente cada tres años con énfasis en diferentes dominios o áreas de evaluación. En el primer estudio (años 2000 y 2002), el énfasis fue lectura; en el segundo estudio (año 2003), fue matemáticas; en el tercer estudio (año 2006), será ciencias.

### Publicación de resultados

Los resultados de la aplicación del año 2000 se difundieron a finales de 2001 en el documento:

- OCDE (2001) *Conocimientos y aptitudes para la vida. Resultados de PISA 2000*. México. Editorial Santillana.

Los resultados de la aplicación complementaria del 2002 (*PISA plus*) se publicaron en julio de 2003 en el documento:

- OCDE (2003) *Aptitudes Básicas para el mundo de mañana. Otros resultados del proyecto PISA 2000*. París. OCDE.

Los resultados de la segunda aplicación (año 2003) serán dados a conocer a finales de 2004

## Coordinación del proyecto

PISA es un proyecto de colaboración internacional que reúne expertos de los países participantes y es monitoreado permanentemente por los gobiernos respectivos. Los órganos creados para su funcionamiento y operación son los siguientes:

- El Consejo de Países Participantes (BPC por sus siglas en inglés). En este órgano, cada país está representado; su misión es determinar, en el contexto de los objetivos de la OCDE, las prioridades de PISA y supervisar permanentemente el programa.
- Los expertos de los países participantes colaboran en grupos de trabajo denominados Grupos de Expertos de PISA, (FEG por sus siglas en inglés) con esto se asegura que los instrumentos son internacionalmente válidos y que toman en cuenta los contextos culturales y educativos de los países participantes. Los grupos de expertos son los siguientes:
  - Grupo Funcional de Expertos en Matemáticas
  - Grupo Funcional de Expertos en Lectura
  - Grupo Funcional de Expertos en Ciencias

Además existe un grupo dedicado especialmente a los aspectos técnicos de la prueba:

- Grupo de Asesoría Técnica
- A nivel nacional (local) el proyecto PISA está a cargo de los Administradores Nacionales del Proyecto (NPM por sus siglas en inglés). Los NPM aseguran que la implementación del proyecto sea de alta calidad y verifican y evalúan los resultados, análisis, reportes y publicaciones.
- Dentro del marco establecido por el Consejo de Países Participantes, el diseño e implementación de los instrumentos del proyecto está a cargo de un consorcio denominado Consortio PISA. El Consorcio es dirigido por el Consejo Australiano para la Investigación Educativa (ACER por sus siglas en inglés). Los otros socios del consorcio son el Instituto Nacional Holandés para la Medición Educativa (CITO), el Instituto Nacional Japonés para la Investigación Educativa (NIER) y el Servicio de Examinación Educativa (ETS) y la empresa WESTAT de Estados Unidos.
- La responsabilidad general de manejo del programa por parte de la OCDE está a cargo del Secretariado de la OCDE, éste actúa como secretario del Consejo de Países Participantes, construye los consensos entre países y sirve como interlocutor entre el Consejo de Países Participantes y el Consorcio Internacional PISA encargado de la implementación del proyecto. El Secretariado de la OCDE también produce los indicadores y análisis, y prepara los reportes internacionales y las publicaciones en cooperación con el Consorcio PISA y en estrecha consulta con los países miembros.

### Órganos de coordinación de PISA

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

### Población objetivo

Con objeto de asegurar la mayor comparabilidad posible, el proyecto PISA se dirige a evaluar jóvenes de 15 años de edad, independientemente del grado y modalidad de enseñanza secundaria o media superior en la que estén inscritos; en la mayoría de los países, en esta edad los estudiantes están finalizando la educación básica obligatoria y se aprestan para ingresar al mundo del trabajo, de manera que es posible apreciar las competencias adquiridas en una etapa semejante de vida.

### Muestra

Se utilizan muestras representativas entre 4,500 y 10,000 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 30,000 estudiantes para poder derivar conclusiones a nivel estatal y no sólo a nivel país.

Sobre la base de un marco muestral de escuelas de cada país, el consorcio encargado de la prueba selecciona de manera aleatoria la muestra de escuelas para la aplicación del estudio, considerando escuelas de reemplazo para las contingencias en las que no sea posible la aplicación. La posibilidad de reemplazar escuelas se limita a un porcentaje ya fijado. Cada país organiza su aplicación, calificación de pruebas y análisis de resultados, con la supervisión de el consorcio y el secretariado de la OCDE.

### **Instrumentos, escalas y procedimientos para su aplicación**

Las pruebas son de papel y estandarizadas para ser llenadas con lápiz; contienen reactivos de respuesta estructurada u opción múltiple así como un porcentaje alto de reactivos o preguntas de respuesta construida o abierta y están diseñadas para ser respondidas con holgura en un máximo de dos horas. El diseño de las pruebas, aplicación y análisis de resultados son controlados por un equipo internacional con estándares de calidad probados. Si bien se delega en cada país la traducción de los instrumentos a la lengua nacional, dicha traducción se verifica y valida por parte del consorcio encargado de la operación de la prueba. Además de las pruebas de conocimientos y habilidades y con objeto de obtener mayor información sobre el contexto en el que se desarrolla el aprendizaje, se aplican cuestionarios de contexto a los estudiantes, a las autoridades de la escuela y al director del plantel.

El diseño del estudio considera una escala continua donde cada estudiante recibe una puntuación que muestra la actividad o conjunto de tareas con mayor grado de dificultad que es realizado correctamente. Cada área de competencia -lectura, matemáticas y ciencias- tiene su propia escala, donde la puntuación promedio es de 500 puntos.

## El proyecto PISA y otras pruebas internacionales

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

Su *origen*: La iniciativa ha sido tomada por los gobiernos y el estudio estará diseñado para servir a su interés especialmente para los temas de política educativa.

Su *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.

La *edad del grupo*: La evaluación de los jóvenes que están acabando el periodo de escolarización obligatoria aporta un indicador muy 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 currículum básico común, en términos generales. En este punto de la formación, resulta útil determinar el grado en que los alumnos han adquirido el conocimiento y las destrezas necesarios que les ayudarán en el futuro para desenvolverse en la vida, incluyendo las opciones específicas de aprendizaje que pueden elegir más tarde.

El *conocimiento y las destrezas evaluados*: Estos 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. Esta 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 currículum, pero los evalúa sobre todo en términos de adquisición de destrezas y conceptos amplios, que 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.

Este énfasis en la evaluación en términos de la adquisición de conceptos amplios resulta especialmente significativo si se tiene en cuenta el interés de los países en cuanto al desarrollo del capital humano que la OCDE define como:

*“el conocimiento, destrezas, competencias y otros atributos ligados a las personas que son relevantes para el bienestar personal, social y económico”.*

Los cálculos del nivel de capital humano, o de la base de las capacidades humanas, han sido obtenidos habitualmente, y en el mejor de los casos, mediante parámetros como el nivel de educación alcanzado. Estos parámetros se muestran aún más como inadecuados cuando el interés en el capital humano se extiende para incluir atributos que permitan una plena participación social y democrática en la vida como adultos y que doten a los alumnos de la preparación necesaria para seguir aprendiendo continuamente a lo largo de la vida.

La evaluación directa del conocimiento y destrezas al final del periodo de educación básica permite al proyecto PISA examinar el grado de preparación de los jóvenes para la vida como adultos y, hasta cierto punto, analizar la efectividad de los sistemas educativos. La meta del proyecto consiste en la evaluación del rendimiento de los sistemas educativos en relación con sus objetivos subyacentes (tal como los define la sociedad) y no en relación con la enseñanza y aprendizaje de un cuerpo de conocimientos.

## 2

### 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.

## 1. Cómo se miden los conocimientos y aptitudes en PISA

Las habilidades que se miden en PISA deben ser consideradas 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 algo arbitrario.

Sin embargo, es útil dividir este continuo en etapas o niveles para comunicar la competencia de los estudiantes en algo más que números. Esto quiere decir que se debe considerar a un estudiante dentro de un nivel si la expectativa que se tiene de él es que es probable que resuelva las tareas de ese nivel más que fallarlas. Esto implica que el estudiante obtenga al menos la mitad de los reactivos de una prueba de ese nivel correctos. De hecho, la esperanza de éxito de un estudiante en un nivel particular en una prueba que contuviera solo reactivos de ese nivel sería de un 50% para un estudiante en la parte en la parte baja del nivel y superior a 50% para los otros estudiantes más arriba en el mismo nivel.

Para facilitar la interpretación de los puntajes asignados a los estudiantes, PISA ha definido 5 niveles de competencia que se expresan en una escala que va de 200 a 800 puntos con un puntaje promedio de 500 puntos y aproximadamente dos tercios de los estudiantes de los países de la OCDE con puntajes entre 400 y 600 puntos (+/- una desviación estándar).

## 1.1 Significado de los niveles de competencia en lectura

En las escalas de lectura el nivel 5 de competencia corresponde a puntajes de 625 o más; el nivel 4 a puntajes en el rango de 553 a 625; el nivel 3 a puntajes de 481 a 552; el nivel 2 a puntajes de 408 a 480, y el nivel 1 a puntajes de 335 a 407. Desde luego, un estudiante en un nivel particular no solo demuestra conocimientos y habilidades en ese nivel sino también las competencias requeridas en los niveles inferiores; es decir, un estudiante en el nivel 3 es también competente en los niveles 1 y 2.

Por otra parte, los estudiantes con puntajes por debajo de los 335 puntos; es decir, que no alcanzan el nivel 1, no muestran las habilidades más básicas que mide PISA.

En el caso de las habilidades para la lectura la descripción de los niveles es como sigue:

### **Domino en el nivel 5 (más de 625 puntos)**

Los estudiantes que dominan el nivel 5 de la escala, 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; ser capaces de evaluar críticamente y establecer hipótesis, recurrir a conocimiento especializado e incluir conceptos que puedan ser contrarios a las expectativas.

### **Dominio en el nivel 4 (entre 553 y 625) puntos)**

Los estudiantes que dominan el nivel 4 de la escala, 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. Para el área de la OCDE en conjunto, 31% de los estudiantes dominan el nivel cuatro o más (es decir, los niveles 4 y 5). La mitad de los estudiantes de Finlandia y 40% o más de los de Australia, Canadá, Irlanda, Nueva Zelanda y el Reino Unido, alcanzan por lo menos el nivel 4. Con excepción de Luxemburgo y México, por lo menos uno de cada 5 estudiantes de los países de la OCDE alcanza como mínimo el nivel 4 o mayor.

### **Dominio en el nivel 3 (entre 481 y 552 puntos)**

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

### **Dominio en el nivel 2 (entre 408 y 480 puntos)**

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

## **Dominio en el nivel 1 (entre 335 y 407 puntos) o menor (menos de 335 puntos)**

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”. Dado que comparativamente pocos adultos jóvenes en los países de la OCDE no cuentan con destrezas técnicas de lectura, PISA no busca medir cosas como el grado al cual los estudiantes de 15 años leen fluidamente, la calidad de su ortografía y su capacidad para reconocer palabras. En línea con los puntos de vista más modernos acerca de la aptitud para la lectura, PISA se concentra en la medición del grado al 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 Nivel 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.

Los estudiantes cuyo desempeño se sitúa por debajo de los 335 puntos, es decir, debajo del Nivel 1, no son capaces de realizar el tipo de lectura más básico que PISA busca medir. Esto no significa que no tengan aptitudes para lectura. De hecho, la mayoría de estos estudiantes pueden probablemente leer en el sentido técnico de la palabra y la mayoría (54 % en promedio en los países de la OCDE) son capaces de resolver con éxito al menos 10 % de los reactivos de lectura que no sean de opción múltiple del PISA 2000 (y 6 % la cuarta parte de estos reactivos). No obstante lo anterior, su patrón de respuestas en la evaluación es tal que se esperaría que resolvieran menos de la mitad de los reactivos de una prueba compuesta únicamente del Nivel 1 y, por lo tanto, mostrar un desempeño por debajo del Nivel 1. Dichos estudiantes tienen serias dificultades para

emplear la aptitud para lectura como una herramienta eficaz para ampliar y aumentar sus conocimientos y destrezas en otras áreas. Los estudiantes con aptitud para lectura menores al Nivel 1 podrían, entonces, estar corriendo el riesgo no sólo de enfrentar dificultades en su transición inicial de la educación hacia el trabajo, sino también de fracasar en beneficiarse de una educación más amplia y de las oportunidades de aprender durante toda su vida.

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

Nivel	Puntaje en la escala de PISA
5	Más de 625
4	553 a 625
3	481 a 552
2	408 a 480
1	335 a 407

## 1.2 Significado de los niveles de competencia en matemáticas y ciencias

Dado que en el año 2000 la evaluación de las aptitudes para matemáticas y ciencias fue más limitada que la evaluación de la aptitud para la lectura, no se intentó definir niveles de dominio como se hizo en lectura; pero si es posible proporcionar una descripción amplia del desempeño en matemáticas y ciencias en diversos puntos de la escala.

- Hacia el nivel más alto, 750 puntos, 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. Los estudiantes en este nivel identifican y aplican conocimientos y herramientas relevantes (a menudo en el contexto de un problema con el que no están familiarizados), emplean la perspicacia para identificar maneras adecuadas de encontrar una solución y muestran otros procesos cognoscitivos de alto nivel tales como la generalización, el razonamiento y la argumentación para explicar y comunicar resultados.
- Hacia los 570 puntos de la escala, los estudiantes son normalmente capaces de interpretar, vincular e integrar distintas representaciones de un problema o diferentes fragmentos de información; de manipular y emplear un modelo dado, que a menudo involucra el uso de álgebra u otras representaciones simbólicas; y verificar y revisar proposiciones o modelos dados. Los estudiantes normalmente trabajan con estrategias, modelos o proposiciones dadas (como mediante el reconocimiento y extrapolación de un patrón) y seleccionan y aplican el conocimiento matemático necesario para resolver un problema que puede requerir de un número pequeño de pasos de procesamiento.

- En el extremo bajo de la escala, alrededor de los 380 puntos, los estudiantes son normalmente capaces de completar solamente un paso de procesamiento que consiste en la reproducción de elementos matemáticos básicos o en aplicar habilidades simples de cálculo. Los estudiantes normalmente reconocen la información a partir del material diagramático o de texto que es familiar y directo y en el cual se proporciona la formulación matemática o ésta es claramente aparente. Cualquier interpretación o razonamiento generalmente involucra el reconocimiento de un solo elemento familiar en un problema. La solución requiere de la aplicación de procedimientos rutinarios en un solo paso de procesamiento.

En la evaluación de PISA, el 5% más apto de los estudiantes alcanzó en promedio 655 puntos en los países de la OCDE, el 10% llegó a 625 puntos y el 25% 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% llegaron a 367 y más del 95% tuvieron 326 puntos.

## 2. Resultados en lectura

De acuerdo con las puntuaciones que los estudiantes obtuvieron por la capacidad mostrada para recuperar, interpretar y reflexionar sobre una lectura, así como evaluar la misma, considerando sus experiencias y conocimientos previos, se otorgó una puntuación tomando en cuenta la dificultad de las tareas o actividades involucradas; estas puntuaciones sirvieron de base para clasificar a los estudiantes en cinco niveles, información que a su vez permitió ubicar a los países de acuerdo con el porcentaje de estudiantes competentes en el nivel 3 o siguientes.

Si bien el 60% de los estudiantes de 15 años de los países de la OCDE son competentes para realizar las actividades correspondientes, su porcentaje varía entre países. Por sí solo ningún indicador traduce adecuadamente el desempeño de los estudiantes en cada país. Algunos países muestran diferencias importantes entre el desempeño de los estudiantes en actividades básicas: identificar e interpretar información y las de mayor complejidad como son las que demandan reflexión y evaluación. En el nivel 5 (superior a 625 puntos) solamente el 10% de estudiantes se mostraron competentes. Hay países con más del 15% de estudiantes en este nivel como Nueva Zelanda, Finlandia y Australia, mientras habría otros con menos del 5% de su población. **México y Brasil, los dos únicos países latinoamericanos que participaron en el estudio del 2000, tienen el 1% de su población en este nivel.**

Pisa enfatiza la capacidad del estudiante para servirse de la lectura como medio para aprender, competencia que supera la mera habilidad de leer o de aprender para leer. Las actividades más simples de lectura asociadas a la habilidad para la lectura conformaron el nivel 1 (por ejemplo, localizar información sencilla o identificar el tema principal de un texto). En este nivel se encuentra el 12% de estudiantes evaluados, y un 6% 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% se ubica en el nivel I o por abajo,** cuando países ubicados en los primeros lugares tienen menos del 5% de sus estudiantes en el mismo.

En el modelo PISA los alumnos del nivel 4 son capaces de resolver tareas complejas de lectura, como localizar información impresa, interpretar significados de matices del idioma y evaluar críticamente un texto; el nivel 3 exige complejidad menor: localizan entre diferentes partes de un texto y logran formular relaciones con experiencias cotidianas; y en el nivel 2 resuelven actividades básicas de localización, inferencia y comprensión.

En promedio, poco menos de las tres cuartas partes de los estudiantes se encuentran en estos tres niveles, pero existen fuertes diferencias entre aquellos países donde más de una cuarta parte de estudiantes se desempeñan por debajo del nivel 2, hasta aquellos países como Corea, donde el 89% se encuentra en los niveles medios, con escasa proporción en los niveles 5 y 1, o Nueva Zelanda donde dos terceras partes se ubican en el nivel medio, pero respecto a la otra tercera parte, en el nivel más alto se encuentran tres veces más estudiantes que los de Corea y en el nivel 1 o abajo más de dos veces que los de Corea. **México tiene un 30% de su población en el nivel 2, un 19% en el nivel 3 y otro 6% en el nivel 4. Sin embargo, debe considerarse que México tiene en el nivel 1 o inferior al 44% de los estudiantes de su muestra.**

En los resultados presentados por PISA existe el reconocimiento explícito del nivel de incertidumbre que acompaña al estudio muestral donde el promedio se mueve en un 95% de confianza. Dado que los promedios de varios países son semejantes, se observan traslapes que impiden precisar con toda seguridad la posición de los países. Calcular la media de los puntajes del estudiante, permitió una lectura sintética del desempeño por país, donde se observan diferencias importantes entre países de hasta 125 puntos entre el desempeño más bajo y el más alto; de igual manera, al interior de algunos de estos países las diferencias de desempeño resultan muy marcadas, pareciendo atinada la relación entre alto promedio en desempeño y bajo índice de desigualdades en los países.

### 3. Resultados en matemáticas

PISA 2000 muestra los desempeños de los estudiantes en matemáticas y en ciencias, sin el nivel de profundidad que se exploró para lectura; la evaluación se realizó en escala simple sin utilizar los niveles de desempeño como fue el caso de lectura. Matemáticas y Ciencias como áreas de conocimiento permiten a los estudiantes desarrollar el contexto necesario para acceder a la información vinculada a la medicina, economía, y medio ambiente, todas ellas importantes para la interpretación de los fenómenos del mundo de hoy.

En el caso de las matemáticas, se consideró relevante que los estudiantes pudieran reconocer e interpretar problemas presentes en la cotidianeidad, trasladar éstos al lenguaje matemático, aplicar las matemáticas en su resolución, comprender los procedimientos o métodos utilizados, interpretar, describir y comunicar los resultados. Las actividades se encuentran graduadas en su dificultad, considerando la complejidad y número de cálculos que intervienen, las conexiones e integraciones de información que deban realizarse; así como las representaciones e interpretaciones que deban formularse.

En este marco, se observa que Japón muestra 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 7 países. **En el otro extremo se encuentran Brasil y México con 330 y 380 puntos respectivamente.**

Como un ejemplo del tipo de actividades presentadas a los estudiantes, véase la siguiente tarea considerada de fácil resolución y a la cual se le asocia una puntuación de 403 puntos. *A partir de una gráfica que muestra la velocidad de un auto de carreras a*

*medida que hace un recorrido alrededor de una pista, los estudiantes contestaron un ejercicio de opción múltiple, donde tenían que indicar en qué lugar de la pista se registraba la velocidad más baja. Solamente fue necesaria una simple observación y el suficiente razonamiento del concepto de cambio, para darse cuenta de que la velocidad más baja se mostraría en el punto más bajo del trazo en la gráfica de velocidad.*

## 4. Resultados en ciencias

En este ámbito, las competencias centrales en el estudio se relacionan con la capacidad para utilizar el conocimiento científico, reconocer preguntas comunes en este campo, identificar la información implicada en las investigaciones científicas, realizar afirmaciones y conclusiones a partir de la información científica, así como la capacidad para describir y comunicar la información de la materia. Como en el caso de matemáticas, también en ciencias se consideraron criterios para ubicar el grado de dificultad de las competencias, en cuyo caso son: la complejidad de los conceptos utilizados, la cantidad de datos proporcionados, la cadena de razonamiento requerido y la precisión exigida para comunicar información.

Un tipo de actividad difícil, registrada por lo mismo con una puntuación alta de 666 puntos es la siguiente: *Se mostró a los estudiantes extractos del diario de un científico del Siglo XIX, una tabla con sus observaciones y un comentario, donde se discutía la muerte posparto, debido a una particular fiebre en una gran proporción de madres, en dos salas de un hospital de maternidad. Los estudiantes tenían que indicar por qué la evidencia no sostenía una creencia contemporánea de que los terremotos causaban la fiebre. En esta tarea era necesario que los estudiantes explicaran la importancia de los diferentes índices de mortalidad en las dos salas.*

Los resultados muestran a Japón y a Corea con los resultados más altos en formación científica, quienes acompañados por otros 9 países se encuentran arriba del 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.**

## 5. Conocimientos y habilidades para la vida en relación con el contexto

Como se refirió en su momento, el estudio de PISA consideró cuestionarios sobre datos personales y gestión escolar, aplicados a los alumnos y a los directores respectivamente; la información proporcionada permitió a los expertos de PISA plantear un conjunto de datos e inferencias que importa mostrar.

En la mayoría de los países, más de la cuarta parte de todos los estudiantes no desea asistir a la escuela; aunque en un rango menor del 20% en México, Dinamarca, Portugal y Suecia, y más del 35% de estudiantes en Bélgica, Canadá, y Francia, junto con 3 países más.

Alrededor de la mitad de los estudiantes muestra una actitud positiva respecto a la lectura, pero se observa una variación entre países. Por ejemplo, en Bélgica y Corea menos de una tercera parte considera divertida la lectura, contra dos terceras partes en Dinamarca, México y Portugal. En Matemáticas, sólo una proporción pequeña de estudiantes percibe la materia como algo relevante para su futuro.

Respecto a cómo aprenden los estudiantes, PISA encuentra que el mejor desempeño se asocia con el control del proceso de aprendizaje (que el estudiante entienda qué es lo que necesita aprender, busque conceptos que no comprende, e información adicional a la básica, confirme si recuerda lo aprendido, y si se trata de lo más relevante).

Existe una relación positiva dentro de cada país entre la lectura por gusto y el desempeño, sin embargo, no necesariamente los países cuyos estudiantes dicen no leer por placer tienen desempeños bajos. Es el caso de Japón, donde la mayoría de estudiantes declara no leer por gusto; y es el caso en sentido contrario de países como Brasil, Letonia, México, Portugal y la Federación Rusa donde menos del 20% reconocieron no leer por gusto, y sin embargo, encontrarse sus estudiantes en niveles bajos de desempeño.

Los resultados por género muestran a las mujeres con mayores competencias para la lectura en todos los países participantes; en matemáticas los hombres obtuvieron mejores resultados en la mitad de los países participantes; en ciencias las mujeres obtuvieron mejores resultados en tres países (Letonia, Federación Rusa y Nueva Zelanda) y los hombres en otros tres (Corea, Austria y Dinamarca). En el caso de México, las mujeres muestran 20 puntos más en lectura, los hombres 11 más en matemáticas y 4 puntos en ciencias. De estos resultados parece derivarse que los países proporcionan un contexto de aprendizaje que beneficia de manera diferente la equidad de géneros.

La relación entre contexto familiar y desempeño muestra que un mejor desempeño de los estudiantes se relaciona con una mayor educación de los padres, interrelación positiva entre los miembros de la familia y presencia de recursos económicos, particularmente bienes culturales como obras literarias en el hogar. También en el terreno del contexto familiar se pudo observar que un mejor desempeño está asociado con el hecho de vivir con ambos padres y haber nacido en el país donde se reside, aunque hubo países como Bélgica, Alemania, Luxemburgo y Holanda donde los hijos de inmigrantes vieron su desempeño escolar afectado.

Los hallazgos de PISA respecto a la capacidad de los planteles para compensar el contexto socioeconómico de las familias, muestra que las diferencias en el desempeño entre escuelas y dentro de escuelas están asociadas con el contexto socioeconómico, sin embargo, la composición socio-económica de la población estudiantil en una escuela es un indicador más fuerte en el desempeño del estudiante que la diferencia del contexto familiar individual; escuelas con ambiente favorable y mejores recursos tienden a tener estudiantes con mayores ventajas educativas.

De acuerdo con lo expresado, más allá de la infraestructura, las escuelas deberán considerar la influencia que sobre el rendimiento tiene el uso que hacen los estudiantes de los recursos escolares; importa tener una planta fuerte de docentes titulados o con buenos niveles de formación y actualización; controlar la cantidad de estudiantes por docente, pues cuando sobrepasan los 25 alumnos el desempeño tiende a bajar. Impactan positivamente aspectos de política y práctica escolar como son expectativas altas respecto a los estudiantes, compromiso docente con su profesión y autonomía escolar; relación profesor-estudiante armoniosa; ambiente de disciplina, así como el énfasis de los profesores en el desempeño académico y el aprendizaje en los estudiantes.

## 6. PISA 2002 “Aptitudes básicas para el mundo de mañana”

### 6.1 Lectura

Los resultados de lectura -que fue el área de conocimiento más destacada en PISA- permiten observar que un 44.2% de estudiantes mexicanos se distribuye en el nivel 1 o por debajo de él, por lo que su situación es de malos lectores; los niveles 2 y 3 atraen a

un 49.1% de alumnos regulares; y el 6.9% de estudiantes se ubica en los niveles 4 y 5 de buenos lectores.

Comparando los resultados de México con los obtenidos por los países participantes de América Latina, se observa que en Argentina el porcentaje de buenos lectores (10.3) supera al de México, sin embargo, el 22.6% de sus estudiantes se ubica por debajo del nivel 1, contra el 16.1% de México. Los otros países de la región presentan una tendencia similar: decrece el número de lectores buenos y se incrementa el de lectores deficientes: 5.3% frente a 48.2% en Chile; 3.7% contra 55.8% en Brasil; 1.1% frente a 79.6% en Perú. Las puntuaciones obtenidas por el conjunto latinoamericano permiten observar a México con 422 puntos, Argentina 418, Brasil 396, Chile 410 y Perú 327 puntos.

Comparando a México con Estados Unidos y Canadá, los dos países americanos miembros también de la OCDE y socios en el Tratado de Libre Comercio, se observa en el primero un 33.7% de lectores con competencia alta y un 17.9% de malos lectores; y en el segundo, un 44.5% de buenos lectores y solamente un 6.9% de lectores deficientes, sus puntuaciones en lectura son 504 y 534 respectivamente. El resto de países muestra la misma situación conforme ascienden en la posición de lugares, Finlandia que ocupa el primer lugar con 546 puntos tiene 50.1% de estudiantes competentes para la lectura, tal y como se concibe en el modelo de PISA.

## 6.2 Matemáticas y Ciencias

Las puntuaciones de los países latinoamericanos son muy similares entre sí, a excepción de Perú, nación que presenta problemas socioeconómicos más agudos (recuérdese que su producto interno bruto per cápita es de sólo 4,799 dólares y el gasto educativo por alumno representa 3,479 dólares; a diferencia de México cuyas cifras son del orden de los 9,117 dólares y 12,189 dólares respectivamente).

En este marco, México toma ligeramente la delantera en las puntuaciones o camina en situación muy pareja respecto a sus pares. En Matemáticas obtiene 387 puntos, Argentina 388, Brasil 334, Chile 384 también y Perú 292. En Ciencias, México obtiene 422 puntos, Argentina 396, Brasil 375, Chile 415 y Perú 333. Los dos países industrializados de alto desarrollo del continente americano, Estados Unidos y Canadá, obtuvieron puntajes promedio en Matemáticas de 493 y 533 respectivamente; y en Ciencias de 499 y 529 cada uno. Obsérvese que a excepción de Canadá, ninguno de los países supera el promedio de 500 puntos de la OCDE, situación particularmente espinosa para Estados Unidos cuya potencia y liderazgo mundial se ha intensificado en los últimos años.

En Matemáticas y Ciencias se observa una recomposición muy interesante de los lugares que ocupan los países mejor posicionados en la OCDE: Hong Kong, Japón y Corea toman la delantera en Matemáticas con 560, 557 y 547 puntos cada uno. En Ciencias podemos ver a Corea, Japón y Hong Kong con 552, 550 y 541 puntos respectivamente. Nueva Zelanda, 3er. Lugar en lectura después de Canadá, vendrá a ocupar el 4° lugar en matemáticas y el 7° en ciencias. Y Finlandia, 1er. Lugar en lectura, se sitúa en el 5° de matemáticas y el 4° de ciencias.

## 7. 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, debieran 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 que se sitúa en el primer sitio: habría un 95% de confianza sobre la corrección de sus respectivos lugares; no es el caso, sin embargo, de Canadá quien estando en el segundo lugar, podría también ocupar el tercero o cuarto.

Los aspectos anteriores (multidimensionalidad de las pruebas y traslapes de las puntuaciones) 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, resulta fácil e impactante decir: *México, penúltimo lugar del mundo*, sin considerar múltiples elementos que condicionan el planteamiento; por ejemplo, no todos los países del mundo fueron evaluados, no todos los países 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 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 que ha salido a la luz pública a propósito de los resultados de PISA 2000 y PISA 2002:

Toda comparación exige la comprensión de las peculiaridades que hacen diferente lo que se pretende comparar. Por lo mismo, las lecturas superficiales de los hallazgos nos colocan en riesgo de distorsionar los propósitos y por tanto, los límites del estudio, soslayar el contexto en el cual se realizan las comparaciones y el lugar que tiene México en el marco regional latinoamericano, y en el ámbito del conjunto de países que intervinieron en el estudio.

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 está asociada a un manejo adecuado de los datos empíricos, respetuoso de 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; ésta es la razón por la que cualquier ejercicio intelectual con pretensiones de comparación tiene la exigencia -particularmente en temas altamente complejos por su multidimensionalidad como es el educativo- de considerar el contexto como el factor determinante para comprender de manera integral el comportamiento de la realidad.

La participación de México en el estudio de PISA ha despertado cuestionamientos en el sentido de que los referentes con los cuales se compara México son impertinentes por tratarse de países muchos de ellos, con situaciones socioeconómicas diferentes a las de México y cuyos estudiantes se encuentran en situación de ventaja respecto a los estudiantes nacionales. Siendo innegable esta situación, el provecho que se pretende extraer de la evaluación es conocer el logro de los estudiantes mexicanos en esas áreas de conocimiento, considerando países con características semejantes como pueden ser los integrantes de la región latinoamericana, países de economía media y países de alto desarrollo. Esta comparación -teniendo siempre presente el contexto- resulta enriquecedora para el autoconocimiento como país y para el reconocimiento de las medidas que urge instrumentar para mejorar el sistema educativo nacional.

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 del logro de los estudiantes. Desde esta perspectiva, los cuestionarios de contexto aplicados en el estudio ofrecen información significativa para abordar comprensivamente los escenarios en los cuales tienen lugar los aprendizajes de los estudiantes. Hablan de los entornos familiares, escolares y áulicos que contribuyen o no a que los alumnos aprendan. Conviene señalar, por tanto, algunos factores sobre los que PISA y prácticamente toda la investigación educativa han llamado la atención.

## 8. Los factores que influyen en la calidad del rendimiento escolar

La situación socioeconómica de los estudiantes en convergencia con los factores que componen la vida de una escuela intervienen de manera relevante en su rendimiento. Las condiciones de vida de los estudiantes mexicanos -inferiores a las de los países más desarrollados- tienen un fuerte peso que se evidencia en los resultados educativos bajos respecto a éstos, pero similares respecto a los otros países latinoamericanos con economías parecidas.

Para ilustrar lo anterior obsérvese el caso de dos países latinoamericanos con una situación relativamente semejante, Argentina en el lugar 35 del estudio con PIB *per cápita* de 12,377 dólares y gasto educativo de 18,839 y Chile, lugar 36 con 9,417 y 17,820 dólares respectivamente. Frente a esta realidad obsérvese el caso de Finlandia y Canadá, 1° y 2° lugares del estudio, con 25,357 y 28,130 dólares *per cápita* cada uno y gasto educativo de 47,854 y 59,808 dólares respectivamente.

Existen excepciones que atraen el interés por profundizar en el contexto que las produce, como es el caso de Luxemburgo, quien presentando un PIB *per cápita* de 48,239 -el más alto de todos los países participantes y casi 6 veces el de México que es de 9,117- ocupa el lugar número 31, sólo 3 lugares más alto que México. Y llama también la atención Estados Unidos, quien teniendo un gasto educativo total por alumno de los más altos de los países participantes (72,119 dólares), ocupe el lugar 16 de la tabla, cuando Corea, con menos de la mitad de ese presupuesto (30,246) se ubica en el lugar número 7.

El componente demográfico también impacta, tanto en el ingreso *per cápita* como en el gasto educativo y la profundidad con la que se atienden problemas de equidad y calidad. Los sistemas educativos más eficaces están atendiendo al 100% de su población de 15 años, o se acercan mucho a la atención total. Finlandia y Estados Unidos, junto con otros países desarrollados lo están haciendo. En el bloque latinoamericano México tiene la cobertura más baja (51.6% del universo de 15 años en la escuela), seguido por Brasil (53.2%), Perú (65.6%), Argentina (76.3%) y Chile (87.3).

En esta línea de ideas es posible desprender otro aspecto que aparece interrogando a la calidad de los sistemas educativos con alta proporción de población fuera de las aulas - aunque el abandono escolar y el no ingreso tienen causas que trascienden al sistema educativo, para enraizar en la situación socioeconómica de las familias y aún en la cuestión étnica.

Siguiendo la hipótesis de que los jóvenes de 15 años que han abandonado la escuela, tendrían resultados promedio inferiores a quienes han logrado permanecer, parece correcto desprender que el rendimiento de los estudiantes de países que presentan esta situación está sobreestimado ¿es posible que México hubiera tenido puntajes más bajos aún si un número mayor de estudiantes estuviera siendo atendido por el sistema educativo nacional? La complejidad de esta realidad obliga a la prudencia cuando de

externar juicios triunfalistas se trata. La relativa superioridad que México pueda evidenciar respecto a otros países latinoamericanos, debe pasar necesariamente por el análisis más fino de las condiciones específicas de los factores internos y externos que estarían impactando en los rendimientos educativos.

Cuando se alude a la influencia que genera el factor socioeconómico en los logros de los estudiantes, no se piensa en un determinismo que reste potencia a otros factores como sería la cultura de una sociedad, espacio en el que el sistema de ideas -donde la religión y las diferentes tradiciones juegan un papel importante- explica con gran fuerza los valores, actitudes y comportamientos específicos respecto a las expectativas en torno a la educación y los esfuerzos que demanda, el estilo de gestión institucional requerido, la manera de entender la normatividad y el respeto a los compromisos contraídos, etcétera.

Las políticas educativas también influyen en el comportamiento global de un sistema educativo, y en el microcosmos que constituye cada escuela, haciendo grandes diferencias entre éstas, como es el caso de México y otros países, aunque los resultados de los estudiantes se muestren homogéneos. Además de las políticas educativas que un país toma, considerando entre múltiples factores, los de tipo económico y político tanto a nivel nacional como internacional, las diferencias entre escuelas están asociadas a la diferenciación social que presenta heterogeneidad étnica, defensa de culturas regionales, inexistencia de currículo nacional o autonomía de estados para desarrollar programas específicos, o conflictos específicos por los que atraviesa determinado país.

## 9. La lucha de México por la equidad con calidad. Lecciones que aporta la evaluación de PISA

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, la equidad y la calidad de las oportunidades que constitucionalmente se encuentran garantizadas para todos los mexicanos.

Por ejemplo, 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 que 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 y 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 ángulos de mirada 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 agrandar las muestras de estudiantes de manera que se puedan 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 tamaño.
- Intensificar la participación en evaluaciones internacionales que 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.

#### Nota

Los resultados de PISA están disponibles en diferentes publicaciones y en su página de Internet [www.oecd.pisa.org](http://www.oecd.pisa.org). Algunas de las publicaciones de PISA se muestran en la bibliografía del presente documento.

# 3

## Metodología para el diseño y aplicación de las pruebas PISA

La medición por medio de pruebas estandarizadas a gran escala a diferentes países, como lo es 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 del PISA se caracteriza por ser integral, sistemática y rigurosa.

La metodología abarca un proceso completo de etapas que parten del diseño de los instrumentos, la definición de los aspectos operativos, el procesamiento de datos y el establecimiento del tipo de resultados. De esta manera, la metodología se convierte en la guía de actuación al definir las normas de diseño, procedimiento y análisis del ciclo del estudio.

El diseño y aplicación del estudio de PISA es responsabilidad de un Consorcio Internacional liderado por el Australian Council for Educational Research (ACER).

La aplicación en cada país está a cargo de los Administradores Nacionales quienes acatan los procedimientos técnicos y operativos definidos por el Consorcio para asegurar la implementación con un alto nivel de calidad.

En esta sección se presentan los aspectos técnicos y metodológicos más importantes del proyecto PISA. Está integrado por la descripción resumida del diseño de instrumentos, los aspectos operativos, el procesamiento de datos y la construcción de escalas.

# 1. Diseño de instrumentos

## 1.1 Diseño de las pruebas

El proceso para la elaboración de las pruebas de evaluación de PISA incluye dos etapas:

- Definición del marco de referencia de los dominios de evaluación, y
- Desarrollo de pruebas

### Definición del marco de referencia de los dominios de evaluación

El desarrollo del marco de referencia para lectura, matemáticas y ciencias estuvo a cargo del Consorcio, a través de la incorporación de los diseñadores de las pruebas y los grupos de expertos con la consulta a los centros nacionales. Este trabajo intenso de discusión y definición del marco tomó ocho meses (enero-septiembre de 1998).

Este primer marco de referencia fue sancionado por el Board of Participating Countries (BPC) y publicado por la OECD en 1999, en *Measuring Student Knowledge and Skills: A New Framework for Assessment* (OECD, 1999).

El marco de referencia en las evaluaciones es el sustento teórico y técnico en un nivel de detalle tal que responde a la pregunta básica sobre qué medir. El desglose claro de lo que se medirá, en concordancia con el objetivo del estudio, es de vital importancia, ya que impacta de manera directa en la validez de los instrumentos de medición.

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

Es preciso aclarar que el marco de referencia puede ser ampliado o modificado, sin detrimento de la estructura básica y siempre en correspondencia con el objetivo y los principios originales para los que fue planteado el estudio.

A continuación se presenta una síntesis del marco de referencia establecido para 2000. Éste fue tomado de los Resultados de PISA 2000<sup>1</sup>

*El objetivo de PISA es obtener medidas de la preparación de los jóvenes adultos de 15 años, quienes han terminado la escolaridad obligatoria, primaria y secundaria, para enfrentar los retos de conocimiento actuales que presentan las sociedades.*

*El tipo de evaluación es de carácter prospectivo y se enfoca hacia la capacidad de los jóvenes para emplear sus conocimientos y competencias al enfrentar los retos que presenta la vida real, más que el grado al cual han logrado dominar un plan de estudios específico. Adicionalmente al desempeño del estudiante, se recaban datos de la familia y los factores institucionales que puedan contribuir a explicar las diferencias en dicho desempeño.*

---

<sup>1</sup> OCDE (2001). Conocimientos y aptitudes para la vida. Resultados de PISA 2000. México: Ed- Santillana.

*El objeto de evaluación de PISA se basa en un modelo dinámico de aprendizaje permanente en donde los nuevos conocimientos y aptitudes necesarios para adaptarse con éxito a un mundo cambiante se obtienen continuamente durante toda la vida.*

*En este sentido, PISA se concentra en los temas que los jóvenes de 15 años necesitarán en el futuro y busca evaluar lo que pueden hacer con lo que se ha aprendido.*

*La aptitud se mide conforme a un patrón continuo, no como algo con lo que un individuo cuenta o no. Se considera que una persona con aptitud cuenta con una gama de competencias. La adquisición de aptitudes es un proceso que dura toda la vida y que se obtiene no sólo a través de la escuela o el aprendizaje formal, sino también mediante la interacción con los iguales y la comunidad.*

*Las aptitudes o áreas de evaluación son Lectura, Matemáticas, Ciencias y Solución de Problemas.*

***Aptitud para la lectura:*** *Capacidad de comprender, emplear y reflexionar sobre textos escritos con el fin de lograr las metas individuales, desarrollar el conocimiento potencial, y participar eficazmente en la sociedad. Comprende reactivos con tres dimensiones particulares.*

***La forma del material de lectura*** *o el texto. Se incluyen textos continuos en prosa y distingue entre distintos tipos de prosa, tales como narración, exposición y argumentación.*

*Además, se incluyen textos discontinuos que presentan información como listas, formatos, gráficas y diagramas.*

*El tipo de reactivos de lectura, se espera que los estudiantes demuestren su dominio en términos de obtener la información, comprender los textos a nivel general, interpretarlos, reflexionar sobre el contenido y la forma de los textos en relación con su propio conocimiento del mundo, y evaluar y argumentar al respecto de su propio punto de vista.*

*El uso para el cual se elaboró el texto, su contexto o situación. Es decir el objetivo de cada texto.*

*La aptitud para matemáticas. Capacidad de identificar, comprender y practicar las matemáticas, así como de hacer juicios bien fundamentados acerca del papel que las matemáticas desempeñan en la vida privada actual y futura de un individuo, su vida laboral, su vida social con parientes y colegas o iguales y su vida como ciudadano constructivo, comprometido y reflexivo. Esta capacidad implica la aplicación de los conocimientos y aptitudes de manera funcional más que su simple dominio en el contexto de un plan de estudios. Los reactivos son de tres tipos:*

*El contenido matemático. Se define en términos de conjuntos de conceptos matemáticos relevantes y conectados que se presentan en situaciones y contextos reales. Éstos incluyen cantidades, espacios, formas, cambios y relaciones e incertidumbre.*

*El proceso matemático. Se incluyen tres conjuntos de competencias, el primero reproducción se compone de cálculos simples o definiciones del tipo más familiar en las evaluaciones convencionales de matemáticas; el segundo conexiones requiere de la reunión de ideas y procedimientos matemáticos para resolver problemas directos y razonablemente familiares; el tercer conjunto reflexión se basa en razonamiento matemático, generalización e introspección, y requiere que los estudiantes realicen análisis, identifiquen los elementos matemáticos en una situación y que planteen sus propios problemas.*

*Las situaciones en las que se emplean las matemáticas. Las aptitudes para matemáticas son evaluadas mediante la asignación de tareas “auténticas” basadas en situaciones que, aunque a veces sean ficticias, representan los tipos de problemas que se encuentran en la vida real. Las situaciones se clasifican como de la vida privada o situaciones personales, de la vida escolar, del trabajo y los deportes, de la comunidad local y la sociedad, y circunstancias científicas.*

***La aptitud para ciencias.** Se relaciona con la capacidad de pensar científicamente en un mundo en el que las ciencias y la tecnología dan forma a la vida. PISA define, además, la aptitud para ciencias como la capacidad de emplear el conocimiento científico para identificar preguntas y extraer conclusiones basadas en evidencias, con el fin de comprender y apoyar la toma de decisiones acerca de la naturaleza y los cambios que se le realizan a través de la actividad humana.*

*PISA incluye también tres dimensiones:*

**Conceptos científicos:** *Los estudiantes requieren dominar una serie de conceptos clave. Se incluyen conceptos de física, química, ciencias biológicas y las ciencias de la tierra y el espacio.*

**Procesos científicos.** *Se evalúa la capacidad para emplear la comprensión y el conocimiento científico y, en particular, la capacidad para adquirir, interpretar y actuar de acuerdo con la evidencia. Se evalúan cinco de estos procesos: el reconocimiento de las preguntas científicas; la identificación de la evidencia; la obtención de conclusiones; la comunicación de estas conclusiones; y la demostración de la comprensión de conceptos científicos.*

**Situaciones científicas y ambiente de aplicación.** *Las preguntas de PISA se agruparon en tres áreas en las cuales se aplica la ciencia: la ciencia en la vida y en la salud; la ciencia en la Tierra y el medio ambiente; y la ciencia en la tecnología.*

*La evaluación se realiza en las escuelas, en una sesión de dos horas, con una prueba escrita, más media hora para responder un cuestionario acerca del propio estudiante. A los directores de escuela se les pide información sobre las características de la escuela en otro cuestionario de media hora de duración.*

*Una característica de las pruebas PISA es que incluye una proporción alta de reactivos de respuesta construida o abierta, dado el tipo de enfoque que maneja para evaluar.*

## Desarrollo de pruebas

La segunda etapa del diseño de las pruebas corresponde en esencia al propio desarrollo de las pruebas. Este es un proceso intenso y riguroso que incluye diversas actividades y tareas.

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.

## Convocatoria para el diseño de reactivos

En tanto que PISA es un estudio comparativo de índole internacional debe asegurarse de allegarse reactivos de diferentes culturas e idiomas. Es en este sentido que el Consorcio convoca a los países participantes a enviar reactivos, de acuerdo con unas guías elaboradas *ex profeso*. Las guías incluyen una lista de variables, tales como: tipo y formatos de texto, formato y contexto de respuesta para que los países puedan clasificar sus reactivos.

## Diseño de reactivos y revisión técnica y de congruencia

El grupo encargado del diseño de las pruebas revisa los reactivos enviados por los países para asegurarse que sean consistentes con el marco de referencia y que técnicamente estén bien diseñados.

En el Consorcio hay un grupo de redactores de reactivos por dominio, cuya responsabilidad es diseñar más reactivos para satisfacer las necesidades de reactivos para

cada área de evaluación. Otra función de los grupos de redactores es la presentación de las intenciones de los reactivos, así como asegurarse de que las traducciones no cambien la intención ni la dificultad de los reactivos.

Durante las traducciones de los reactivos, los redactores también proporcionan sugerencias de adaptaciones a los reactivos, de tal forma que los cambios no alteren lo que los reactivos están midiendo.

### **Revisión de reactivos por un grupo de expertos**

Los reactivos son sometidos a un proceso de revisión por parte de expertos que incluye a los diseñadores de la prueba, revisores nacionales de cada país y por parte de los FEGs.

El proceso de revisión inicia con la revisión de expertos (miembros del staff de PISA y otros no involucrados en el proyecto), después los reactivos se envían a los países participantes para una revisión nacional, y finalmente los FEGs de cada dominio los revisa.

### **Prepiloteo y traducciones**

Se lleva a cabo un prepiloteo para probar los reactivos y de esta manera detectar su claridad y el tiempo de respuesta para cada reactivo (en promedio se requiere de dos minutos por reactivo para responder), además se recolectan respuestas tipo para construir ejemplos de respuesta que puedan incluirse en las Guías de Calificación. Para el prepiloteo deben responder en promedio 35 estudiantes cada reactivo.

Las traducciones de los materiales de evaluación se realizan del inglés al francés y viceversa, a fin de tener dos versiones y que los grupos de trabajo señalen errores tipográficos, ambigüedades, dificultades de traducción y aspectos culturales. Los reactivos traducidos son revisados por los miembros del FEG de lectura para asegurarse de la precisión lingüística.

## 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.

## Prueba piloto, selección, revisión técnica y de expertos de los reactivos para el estudio; y revisión cultural (cultural review panel)

En promedio cada reactivo es respondido por 150 ó 200 estudiantes por país. La revisión y selección de los reactivos para la prueba definitiva se realiza por parte de los revisores nacionales, el grupo de diseño de la prueba, los FEGs y el grupo de expertos encargados de la revisión cultural de los reactivos.

Después de la prueba piloto, se realizan las siguientes actividades:

- A cada país se le envía un formato de revisión de reactivos para ser llenado por el NPM.
- Los diseñadores de la prueba se reúnen para: a) identificar reactivos inapropiados para el estudio, b) seleccionar los mejores reactivos, y c) asegurar el balance de los reactivos conforme al marco de referencia.
- Los FEGs y los diseñadores de la prueba revisan los reactivos seleccionados y presentan una propuesta de reactivos a los NPMs.
- Los reactivos son revisados por el panel cultural para asegurarse que los reactivos seleccionados son adecuados para las diferentes culturas y están libres de sesgo. La adecuación cultural se define como el nivel en que los reactivos son convenientes para la población de 15 años de todos los países participantes, que los contenidos sean familiares a los estudiantes y que las pruebas no violen algún

valor o aproximación cultural. Este grupo también revisa aspectos de género y la adecuación a diferentes grupos socioeconómicos.

- El Consorcio revisa las observaciones que se envían de cada país sobre los materiales de evaluación.
- Se prepara un reporte con los reactivos problemáticos como los que resultaron más fáciles o difíciles de lo esperado, con puntos biserials positivos o negativos, etc. Algunos de estos problemas se deben a la traducción o errores de traducción, pero en otros casos los datos estadísticos señalan que los reactivos tienen errores sistemáticos que es necesario considerar.
- El estudio de confiabilidad de la calificación provee información sobre el acuerdo entre calificadores y se emplea para revisar los reactivos y decidir si los reactivos se mantienen o eliminan, así como para la Guía de Calificación.
- El estudio del funcionamiento diferencial de los reactivos (Differential Item Functioning) analiza el género, la competencia para la lectura y el nivel socioeconómico. Con los resultados de este estudio se define si los reactivos se mantienen o se desechan.
- Los diseñadores de la prueba monitorean el porcentaje de reactivos de cada categoría en cuanto a ciertos atributos de la respuesta.

### **La conformación del banco de reactivos e integración de pruebas**

El banco de reactivos de las pruebas de PISA 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. Los reactivos son auditados con regularidad para asegurar que los instrumentos se desarrollen de acuerdo con las especificaciones del marco de referencia.

Una vez que los reactivos y los materiales son revisados y ajustados, se prepara para ser enviado a los países para la aplicación, de acuerdo con las fechas programadas en cada país.

Para la evaluación del 2000 se diseñaron nueve cuadernillos, cuya conformación por área de evaluación y distribución de módulos se presenta en los cuadros 1 y 2.

**Cuadro 1. Conformación de las áreas de evaluación 2000**

ÁREA	NÚMERO DE REACTIVOS	NÚMERO DE MÓDULOS
Lectura	141	9
Matemáticas	32	4
Ciencias	35	4
<b>Total</b>	<b>208</b>	<b>17</b>

**Cuadro 2. Distribución de módulos por cuadernillo 2000**

Cuadernillo	Bloque 1	Bloque 2	Bloque 3	Bloque 4
1	R1	R2	R4	M1, M2
2	R2	R3	R5	S1, S2
3	R3	R4	R6	M3, M4
4	R4	R5	R7	S3, S4
5	R5	R6	R1	M2, M3
6	R6	R7	R2	S2, S3
7	R7	R1	R3	R8
8	M4, M2	S1, S3	R8	R9
9	S4, S2	M1, M3	R9	R8

En la evaluación del 2003, se diseñaron 13 cuadernillos y se integró un área más de evaluación: solución de problemas. La conformación por área de evaluación y distribución de módulos se presenta en los cuadros 3 y 4.

Cuadro 3. Conformación de las áreas de evaluación 2003

ÁREA	NÚMERO DE REACTIVOS	NÚMERO DE MÓDULOS
Matemáticas	86	7
Solución de problemas	19	2
Lectura	32	2
Ciencias	35	2
<b>Total</b>	<b>172</b>	<b>13</b>

Cuadro 4. Distribución de módulos por cuadernillo 2003

Cuadernillo	Bloque 1	Bloque 2	Bloque 3	Bloque 4
1	M1	M2	M4	R1
2	M2	M3	M5	R2
3	M3	M4	M6	PS1
4	M4	M5	M7	PS2
5	M5	M6	S1	M1
6	M6	M7	S2	M2
7	M7	S1	R1	M3
8	S1	S2	R2	M4
9	S2	R1	PS1	M5
10	R1	R2	PS2	M6
11	R2	PS1	M1	M7
12	PS1	PS2	M2	S1
13	PS2	M1	M3	S2

## 1.2 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:

- Recolección de preguntas o instrumentos usados en otros estudios,
- Desarrollo del marco de referencia, y
- Diseño de los instrumentos

### Recolección de preguntas o instrumentos usados en otros estudios

Se buscaron instrumentos internacionales que pudieran ser fuente de preguntas. Tanto los expertos como los miembros del Consorcio emprendieron la tarea de diseñar preguntas, de tal forma que el conjunto de preguntas elaboradas se sometió a una revisión de expertos sobre la base de:

- Qué tan bien las preguntas miden las dimensiones propuestas,
- Alcance de los conceptos,
- Facilidad de traducción,
- Facilidad de respuesta,
- Adecuación cultural, y
- Validez de *facie* o inicial.

## Definición del marco de referencia

El trabajo de Travers y Westbury (1989)<sup>2</sup> es la base teórica que sustenta los cuestionarios de contexto de PISA. El cuadro 5 tomado del trabajo de Travers y Westbury es la referencia para PISA en el diseño de los cuestionarios de contexto.

**Cuadro 5. Mapeo de la cobertura de los cuestionarios de contexto 2000**

	Antecedentes	Contexto	Contenido
Sistema	1. Rasgos del país	2. Instituciones y políticas	3. Resultados escolares esperados
Escuela	4. Características de la comunidad y la escuela	5. Condiciones y procesos escolares	6. Currículum implementado
Clases	7. Características de los profesores	8. Condiciones y procesos de las clases	9. Currículum implementado
Estudiante	10. Características de los estudiantes	11. Conductas de los estudiantes en el salón de clases	12. Logros escolares

Los factores que influyen el aprendizaje de los estudiantes en cualquier país se consideran contextuales. El contexto es el conjunto de circunstancias bajo las que un estudiante aprende y tiene antecedentes que lo define de una manera fundamental. Los antecedentes provienen de procesos históricos y de desarrollo. Un contexto tiene diferentes expresiones con diferentes contenidos. Para PISA el contenido se basa en un modelo dinámico de aprendizaje para toda la vida en donde los conocimientos y habilidades de adaptación al mundo cambiante se van adquiriendo continuamente.

<sup>2</sup> Travers, K. J. and Westbury, I. (eds.) (1989). The IEA Study of Mathematics I: Analysis of Mathematics Curricula. Oxford: Pergamon Press.

Los antecedentes, el contexto y el contenido representan una dimensión del marco de referencia. Cada uno encuentra expresiones en diferentes niveles: en el sistema educativo, la escuela, el salón de clases y en el estudiante. Estos cuatro niveles constituyen la segunda dimensión del marco de referencia. El texto marcado en negritas del Cuadro 5 corresponde a los elementos que PISA considera importantes en su diseño. El diseño de PISA permite obtener mucha información de la escuela y el estudiante, pero no es el caso para la clase.

### Diseño de los cuestionarios

El diseño de los cuestionarios comprende un proceso de revisión, evaluación, consulta, prueba piloto, entrevistas a estudiantes y el desarrollo de la estructura de los cuestionarios. A fin de validar las preguntas, se realiza un estudio en los países participantes para determinar lo apropiado de las preguntas. Se revisan las preguntas para minimizar los malentendidos o falsas interpretaciones basadas en los diferentes sistemas educativos, culturas u otros factores sociales.

En el Cuadro 6 se presentan las variables y contenido de los dos cuestionarios de contexto.

Se diseñaron dos cuestionarios adicionales dirigidos a los estudiantes, uno sobre las prácticas de aprendizaje y estudio; y otro sobre la familiaridad con las computadoras.

En la evaluación del 2000 y del 2003 en general se mantuvieron los mismos contenidos para los instrumentos de contexto.

Cuadro 6. Variables y contenido de los cuestionarios de contexto 2000

CUESTIONARIO DEL ESTUDIANTE	
Variable	Contenido
Datos demográficos	Fecha de nacimiento, grado escolar, género, estructura familiar, número de hermanos y orden en el nacimiento.
Antecedentes familiares y nivel socioeconómico	Incorporación de los padres al mercado laboral, ocupación de los padres, educación de los padres, país de nacimiento de los padres y el estudiante, idioma hablado en la casa, comunicación social y cultural con los padres, sostenimiento familiar, actividades relacionadas a la cultura, salud de la familia, recursos educativos en casa, uso de recursos escolares, artículos de casa relacionados con la cultura.
Descripción del estudiante de los procesos escolares/instruccionales	Tamaño de la clase, disciplina, relación alumno/profesor, presión de logro, apoyo docente, frecuencia de tareas en casa, tiempo para las tareas en casa, tiempo dedicado a clases, sentido de pertenencia, asistencia a clases remediales, calificaciones en matemáticas, ciencias y español.
Actitudes del estudiante hacia la lectura, las matemáticas o las ciencias	Hábitos, tiempo dedicado, número de libros en casa. (El contenido se modifica, dependiendo del énfasis que se dé a los dominios de evaluación, ya sea lectura, matemáticas o ciencias).
Acceso del estudiante a recursos educativos fuera de la escuela	Recursos disponibles en casa, como: libros, computadora y un lugar de estudio.
Esquemas institucionales de participación y programas de orientación	
Carrera y expectativas educacionales	Expectativas de carrera y de empleo.
CUESTIONARIO PARA LA ESCUELA	
Características de la escuela	Al director se le pide información sobre: localización, tamaño, horas y tipo de la escuela; estructura institucional, niveles y programas escolares ofrecidos, políticas de admisión y transferencia, periodos escolares.
Políticas y prácticas escolares	Centralización o descentralización de la escuela, autonomía y toma de decisiones docentes, prácticas de evaluación, involucramiento de los padres.
Ambiente escolar	Percepción del profesor y del estudiante de factores que afectan el ambiente escolar, percepción de la moral y el compromiso de los profesores.
Recursos escolares	Calidad de los recursos, infraestructura física, tasas de profesores/estudiantes, disponibilidad de computadoras, carencia de profesores.

## 2. Aspectos operativos

### 2.1 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. La definición operacional de la edad de la población depende de las fechas de la aplicación. El requisito internacional es que las evaluaciones se realicen durante un cierto periodo, llamado “ventana de evaluación”, además no deben realizarse durante los tres primeros meses del año escolar.

Los estándares de calidad que PISA marca para el muestreo se refieren a:

- La cobertura de la población objetivo,
- La precisión y confiabilidad, y
- Las tasas mínimas de respuesta de escuelas y estudiantes.

En un estudio internacional se debe asegurar la mayor cobertura de la población a quien se dirige el estudio. Las exclusiones deben estar claramente definidas y tener tasas de exclusión mínimas para asegurar la representatividad. Las exclusiones pueden ser de escuelas y de alumnos; en cualquier caso PISA establece que la tasa de exclusiones sea por debajo del 5%.

Con el propósito de conseguir un alto nivel de confiabilidad y precisión se requiere de un mínimo de 150 escuelas seleccionadas por país. Dentro de cada escuela participante, se necesitan 35 estudiantes seleccionados al azar. En total, PISA demanda un tamaño de muestra mínimo de 4,500 estudiantes.

Otros requisitos de PISA se refieren a las tasas de respuesta de escuelas y estudiantes. Para las escuelas debe ser de 85% y para los estudiantes de 80%.

## Muestreo de escuelas

El diseño de la muestra de escuelas debe ser ordenado o estratificado y seleccionado sistemáticamente al azar de una lista completa de escuelas elegibles de cada país (marco muestral).

Todos los NPMs deben construir un marco muestral de escuelas que corresponda a la población blanco definida. Este marco muestral debe tener una cobertura completa de escuelas que abarque a la población objetivo que demanda el estudio. La calidad del marco muestral incide en los resultados del estudio, ya que afecta de manera directa la probabilidad de selección de escuelas y sus ponderaciones, así como en las estimaciones del estudio.

Antes de muestrear, las escuelas se deben ordenar o estratificar al interior del marco muestral.

La estratificación consiste en clasificar las escuelas en grupos, de acuerdo con ciertas variables de estratificación, tales como: tipo de escuela, tamaño de la escuela, grados escolares, etc.

Las razones para usar la estratificación son:

- Mejorar la eficiencia del diseño de la muestra y en consecuencia lograr que las estimaciones del estudio sean más confiables,
- Aplicar diferentes diseños de muestra, como muestras desproporcionadas, a grupos específicos de escuelas pertenecientes a estados, provincias u otras regiones.

- Asegurar que todos los elementos de la población sean incluidos en la muestra, y
- Asegurar una adecuada representación de grupos específicos de la población objetivo en la muestra final.

Hay dos tipos de estratificación: explícita e implícita. La primera consiste en construir listas separadas de escuelas conforme a las variables de estratificación a considerar. La implícita consiste en ordenar las escuelas dentro de cada estrato por medio de un conjunto de variables implícitas de estratificación. Este tipo de estratificación es una manera muy simple de asegurar una muestra proporcional de escuelas a través de estratos implícitos.

A cada muestra de escuelas se le asignan dos reemplazos tomados del marco muestral. La escuela que inmediatamente sigue a la escuela muestreada se le asigna como reemplazo 1, y la escuela que inmediatamente le precede a la escuela muestreada es el reemplazo 2.

### **Muestreo de estudiantes**

El muestreo de estudiantes se lleva a cabo en cada país a partir del muestreo de escuelas aprobado por PISA. Cada escuela de la muestra debe enviar la lista completa de estudiantes de 15 años al NPM.

Una vez que el NPM recibe las listas de estudiantes elegibles de las escuelas, se realiza el muestreo de estudiantes mediante el uso de un software específico (KeyQuest).

Los estudiantes elegidos de cada escuela son listados en un formato especificado por PISA (Student Tracking Form), éste se considera como el documento central para la administración de las pruebas. Estos formatos son enviados a las escuelas para verificar los datos de los estudiantes y completar los datos que faltan, como la participación y los alumnos con necesidades especiales de educación que participan en el estudio.

## 2.2 Traducción y adecuación de las pruebas y del material del estudio

La causa principal de un pobre comportamiento de los reactivos en las pruebas internacionales son los errores de traducción.

Para PISA que ha desarrollado escalas descriptivas, los errores de traducción son una gran preocupación, pues su interpretación puede ser severamente sesgada por la inestabilidad que presentan los reactivos de un país a otro.

Por ello, PISA ha desarrollado estrictos procedimientos de traducción y equivalencia que son evaluados antes de administrar la prueba.

Algunos de los procedimientos y resultados instrumentados por PISA para asegurar la equivalencia en la traducción y adecuación cultural de la evaluación son los siguientes:

- Proporciona a cada país dos versiones de cada material, una en inglés y otra en francés para que se traduzcan al idioma nacional de manera independiente y luego, a partir de las traducciones, se concilie una tercera versión que será la nacional.
- Sistemáticamente se añade información a los materiales de prueba y al cuestionario para ser traducida. Una vez hecha la traducción se les anexa “Notas de Traducción”, por posibles problemas de adaptación o traducción.
- Una de las partes más importantes del Manual PISA del Administrador del Proyecto Nacional, es precisamente la detallada traducción – adaptación de las guías para la prueba y su revisión después de la prueba de campo.

- Se entrena al equipo principal del proyecto de cada país, en torno a los procedimientos recomendados de traducción.
- Nombra y entrena un grupo internacional de verificadores, que son traductores profesionales expertos en inglés y francés con dominio en cada idioma objetivo. Este equipo verifica las versiones traducidas por cada país y las compara contra las versiones fuente.

### **Doble traducción proveniente de dos idiomas fuente**

La verificación de las equivalencias entre los idiomas “fuente” (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.

Verificación internacional de las versiones nacionales

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 de inglés y francés.

Existen dos centros coordinadores establecidos, uno en el Australian Council for Educational Research (ACER) en Melbourne, el segundo en cApStAn en Bruselas.

Las verificaciones que realiza el Consorcio son para aquellas versiones que atienden a más del 5% de la población objetivo.

Por regla general, los mismos verificadores son usados para versiones del mismo idioma, sin embargo, el español de España no es igual al de México, lo mismo que el portugués de Portugal no es el mismo de Brasil, por lo cual para estos cuatro países se designan traductores nativos independientes.

## Cualidades psicométricas de las versiones nacionales

En los países en donde el idioma es diferente al inglés y el francés, las versiones de PISA y los instrumentos para el proceso de campo fueron a través de los siguientes procedimientos recomendados:

- Doble traducción del inglés, con revisiones cruzadas contra el francés
- Doble traducción del inglés, sin revisión cruzada contra el francés.
- Traducción sólo del inglés o del francés
- Una mezcla de métodos

La variación de resultados en el número de reactivos defectuosos entre cada grupo de países que utilizaron el mismo método muestra que, dicho método no es el único factor determinante de la calidad psicométrica alcanzada en el desarrollo de los instrumentos (otro factor significativo puede ser la precisión de los traductores, de los conciliadores o de los verificadores internacionales).

Los resultados confirmaron la hipótesis de que el método más recomendable es el de doble traducción del inglés y del francés, aunque éste no es significativamente diferente en reactivos defectuosos del método en el que se realizó la traducción de un solo idioma fuente.

La doble traducción de un solo idioma también parece efectiva pero sólo cuando se realizan junto con exhaustivas revisiones cruzadas contra otro idioma fuente.

El promedio de longitud de palabras y oraciones difiere de idioma a idioma y obviamente esto no permite el control total de los traductores cuando se trata de adaptar instrumentos, lo cual hace imposible realizar versiones totalmente equivalentes cuando se trata de pruebas con grandes cantidades de material escrito.

Aunque hay idiomas más “largos”, frecuentemente esto es debido a que son más redundantes morfológica o sintácticamente y en dichos casos, la carga que significa la lectura de la prueba puede ser compensada en parte, en el caso de PISA, por no tener requisitos de velocidad (no es una prueba contra reloj).

De los resultados obtenidos de las traducciones hechas a partir de los dos idiomas fuente para las versiones nacionales, sólo 9 participantes han tenido más de un 10% de reactivos defectuosos.

## 2.3 Operaciones de campo

El estudio es instrumentado en cada país según los procedimientos preparados por el Consorcio.

La realización de estas operaciones es coordinada por el Administrador Nacional del Proyecto (NPM), quien a su vez es apoyado por coordinadores en las escuelas y por los aplicadores de los materiales de evaluación.

### Los Administradores Nacionales del Proyecto (NPM)

Son los responsables de instrumentar el proyecto dentro de sus países. Seleccionan la muestra de escuelas asegurando la cooperación de dichos planteles educativos; luego se define la muestra de sustentantes de entre los estudiantes elegibles según las listas que las escuelas proporcionen a través de un formato específico para tal efecto (Ver apartado 2.1).

El NPM debe seguir estrictamente los procedimientos internacionales para seleccionar tanto la muestra de escuelas como la lista de estudiantes para poder presentar los reportes del proceso exigidos por el Consorcio. En caso de no poder hacerlo, debe

enviar al propio Consorcio las listas completas de escuelas y estudiantes elegibles por su edad, para que sea éste quien realice el proceso de selección de las muestras.

Además de esta actividad, el NPM tiene otras responsabilidades operacionales, las cuales incluyen:

- Traducir y adaptar los instrumentos de la prueba, los manuales y los demás materiales de prueba.
- Elegir y entrenar a los aplicadores de la prueba en cada escuela seleccionada
- Calendarizar las sesiones de aplicación y distribuir a los aplicadores.
- Enviar al Consorcio los documentos traducidos para su revisión y aprobación.
- Armar los cuadernillos de prueba de acuerdo con el diseño y distribución especificado por el Consorcio.
- Supervisar la impresión de los cuadernillos y los cuestionarios
- Mantenerse en comunicación con los Monitores de Calidad en las Escuelas nombrados por el Consorcio.
- Coordinar las actividades de los aplicadores y de los monitores de calidad
- Supervisar el embalado y el transporte de todos los materiales.
- Supervisar la recepción de las pruebas y demás materiales en las escuelas que corresponda y cumplimentar la base de datos llamada Data Entry.
- Enviar los materiales que requiere el Consorcio.
- Preparar el informe del NPM y presentarlo al Consorcio

## Coordinadores de Escuela (SCs)

Se encarga de mantener la relación entre el centro nacional del proyecto y los aplicadores, su primera tarea es preparar la lista de estudiantes elegibles y enviarla al NPM.

Antes de la prueba, los SCs debe:

- De común acuerdo con el NPM, establecer la fecha y la hora de la aplicación.
- Recibir la lista de estudiantes seleccionados según la muestra realizada por el NPM y si fuera necesario actualizarla, al identificar estudiantes con limitaciones o incapacidades que les impidan enfrentar la prueba de acuerdo con los criterios establecidos por el Consorcio.
- Recibir, distribuir y recoger el cuestionario de la escuela y entregarlo al aplicador.
- Informar a los profesores de la escuela, a los estudiantes y a sus padres, la naturaleza de la prueba y la fecha en que se realizará y conseguirá, de ser necesario, el permiso de los padres.
- Informar al NPM o al aplicador sobre cualquier cambio de fecha u horario.
- Asistir al aplicador con los preparativos del salón en el que se llevará a cabo la aplicación.

## Los Aplicadores (TAs)

La actividad principal de los aplicadores es administrar la prueba con imparcialidad y uniformemente, en concordancia con los estándares establecidos por PISA.

Para mantener la imparcialidad, el aplicador no puede ser el maestro de matemáticas o ciencias y de preferencia no debe ser parte del equipo de la escuela.

Antes de la fecha de aplicación, los TAs deben ser entrenados en el centro nacional.

Dicho entrenamiento incluye la revisión del Manual del Aplicador de la Prueba así como las actividades y la conducta que deberá seguir durante la aplicación de la prueba y el cuestionario.

Las responsabilidades adicionales del aplicador son las siguientes:

- Recibir el material de evaluación del NPM y encargarse de su seguridad.
- Prestar total cooperación al SC que le corresponda.
- Contactar al SC una o dos semanas antes de la aplicación para confirmar la agenda de trabajo y los planes en general.
- Conducir los arreglos de la sesión de aplicación en coordinación con el SC
- Completar los formatos específicamente diseñados para él y el reporte de la sesión de evaluación, esta última diseñada para reportar los tiempos, la asistencia o los posibles incidentes durante la sesión de aplicación, etc.
- Asegurar que el número de cuadernillos y cuestionarios entregados por los estudiantes concuerden con el número enviado a la escuela.
- Solicitar al SC el cuestionario de la escuela.
- Después de la aplicación de la prueba y los cuestionarios, enviar al NPM todos los materiales.

## Documentación

El NPM es dotado de los manuales que especifican los principales componentes del proceso de evaluación:

- El Manual de Administrador Nacional del Proyecto, en cual especifica detalladamente sus deberes y responsabilidades. Incluye información general sobre PISA; las operaciones de campo; las responsabilidades y tareas de los SCs y los TA; la traducción de manuales e instrumentos de prueba; obtención de la muestra de

estudiantes; ensamblado y envío de materiales; calificación y captura de datos y; documentación que debe ser presentada al Consorcio.

- Manual del Coordinador de Escuela, en el que se consideran al detalle todas las actividades y responsabilidades descritas arriba.
- Manual del Aplicador, el cual no sólo comprende la descripción de los deberes y responsabilidades del aplicador, sino que además se incluye el guión que deberá leer durante la prueba, así como la forma de reintegrar los materiales.
- Manual de Selección de la Muestra, mediante este documento se detallan los pasos que deben seguirse para la selección de las escuelas y estudiantes que habrán de constituir la muestra, así como los reportes que deberán hacerse.

Adicionalmente, en cada manual se incluyen listas de verificación y tablas de calendarización para facilitar las tareas.

El NPM, por otra parte, deberá incluir información propia del país en los manuales del Coordinador y del Aplicador.

Se incluye también una copia de un programa de cómputo (Key Quest), el cual fue preparado por el Consorcio para realizar la selección de la muestra de estudiantes a partir de la lista que las escuelas envían.

En caso de que el NPM desee usar otro programa para seleccionar la muestra, este debe ser aprobado por el Consorcio.

## Procesamiento de la prueba y los cuestionarios

Toda vez que aproximadamente un tercio del dominio de matemáticas y ciencias y más del 40% de los reactivos de lectura requieren que el estudiante escriba su respuesta, es decir son preguntas abiertas, deben ser evaluadas y codificadas a mano por profesionales de las materias.

Este proceso de evaluación y codificación es una compleja operación, crucial para la realización de los procesos de comparación de resultados entre países, por lo que requiere uniformidad en el trabajo de evaluación y codificación que deben realizar los codificadores.

Para hacer comprensibles los criterios para la codificación, durante el proceso de capacitación de los codificadores se incluyen muchos ejemplos de respuestas aceptables y no aceptables, lo cual se hace para cada dominio cognoscitivo de la prueba a través de las Guías de Calificación y Manuales de Capacitación que PISA le proporciona al NPM.

#### Pasos del proceso de codificación

Para poner en marcha el proceso de codificación, el NPM debe llevar a cabo los siguientes pasos:

- Traducir las Guías de Calificación y los Manuales de Capacitación para cada dominio.
- Contratar y capacitar codificadores.
- Seleccionar ejemplos locales de respuestas para usar en el proceso de capacitación de los codificadores.
- Organizar los cuadernillos que han sido regresados de las escuelas
- Hacer la selección y separar los cuadernillos que serán sometidos al proceso de codificación múltiple.
- Organizar los trabajos de codificación Individual de acuerdo con el diseño internacional.
- Organizar los trabajos de codificación múltiple de acuerdo al procedimiento establecido.
- Presentar la submuestra de cuadernillos para el Estudio de Clasificación de Confiabilidad entre Países.

## 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 equipo que realizará la codificación individual y la múltiple.

Algunas de las características que deben reunir los codificadores son las siguientes:

- Tener buen conocimiento de los contenidos de matemáticas, ciencias y lectura del nivel escolar de los estudiantes seleccionados.
- Estar familiarizados con el lenguaje en que se expresan los estudiantes del nivel escolar examinado.
- Comprometerse a trabajar como calificadores durante todo el tiempo que requiera el proceso de codificación, el cual se espera que no tome más de dos meses.

Como apoyo al NPM el consorcio proporciona material para reclutamiento. Este material es similar a las Guías de Calificación pero mucho más breves. Con este material el mismo candidato puede valorar sus capacidades para enfrentar la tarea de calificador.

El número de codificadores requeridos está definido por el diseño del procedimiento de la codificación múltiple y por la muestra seleccionada. Para el estudio principal, en el caso

de una muestra de 4,500 estudiantes, los codificadores requeridos deben ser múltiplos de 8.

Cuando la muestra es mayor, el NPM debe preparar el diseño de de marcadores y enviarlo al Consorcio para su aprobación.

Ya que se requieren varias semanas para completar la tarea de codificación, se recomienda que sean capacitados un número mayor de codificadores para cada dominio, los cuales deben ser considerados como reserva en caso de fallas de alguno de los codificadores.

Como ya se había mencionado, el proceso de codificación múltiple es complejo, por lo que se requiere tener supervisores de tiempo completo. Estos supervisores deben estar familiarizados con los aspectos logísticos, los procedimientos de la codificación múltiple, los procedimientos de revisión de la confiabilidad de la codificación, la agenda del proceso de codificación, así como el contenido de la prueba y las Guías de Calificación. Los supervisores serán además los encargados de hacer las consultas que fuesen necesarias directamente al Consorcio.

También se requieren designar líderes de mesa durante la codificación, cuya tarea principal es monitorear la consistencia en el trabajo codificación realizado por los codificadores. Otra actividad de este personal es apoyar con la rotación de los cuadernillos, así como resolver preguntas en torno a las Guías de Calificación. Ellos a su vez consultarán a los supervisores sobre casos específicos.

Los líderes de mesa deben participar en la codificación, pero también deben dedicar tiempo a monitorear la consistencia.

Adicionalmente son necesarias varias personas para armar en paquetes los cuadernillos que serán codificados de manera múltiple, bajo el diseño establecido para esta forma de codificación.

## Formatos de confidencialidad

Antes de ver o recibir cualquier material de la prueba PISA, los codificadores deben firmar un formato en el que se comprometen a mantener la confidencialidad de los materiales que deberán conocer y trabajar, por lo que les queda prohibida la extracción o difusión de cualquiera de dichos materiales.

## Capacitación Nacional

Todos los codificadores deben ser capacitados sin importar si han tenido experiencia relacionada con la codificación, o hubiesen estado involucrados en el campo específico de la codificación de la prueba PISA.

Para esta capacitación el Consorcio preparó materiales de apoyo para el NPM, dirigidos a cada dominio específico adicionales a las Guías de Calificación.

En la capacitación, el conocimiento de la materia de que se trata en cada dominio es importante, así como los procedimientos de codificación.

Las sesiones de codificación deben estar organizadas en función de cómo debe ser realizada la codificación.

La forma de codificación debe ser reactivo por reactivo, es decir se debe realizar la codificación del mismo reactivo en todos los cuadernillos de cada codificador antes de pasar al siguiente.

La capacitación debe realizarse por cada dominio y de estos cada Módulo en el que esté constituido dicho dominio.

Se debe revisar la Guía de Calificación igualmente por módulos de cada dominio y pedir a los codificadores que codifiquen ejemplos de respuestas, luego de lo cual debe haber un tiempo para revisar los criterios de codificadores adoptados por el grupo, dando tiempo a la discusión y clarificación de dudas y preguntas.

Debe informarse a los codificadores desde el principio, que se requiere de un alto nivel de consistencia en su codificación y que la confiabilidad será frecuentemente revisada por los líderes de mesa y por los supervisores.

Idealmente los líderes de mesa deben ser entrenados antes que los demás codificadores para lograr una mayor familiaridad con los reactivos y las Guías de Calificación. Esto además con el propósito de que ellos apoyen la capacitación y que en el trabajo de cada mesa ayuden a alcanzar consenso en los criterios de codificación.

Estos líderes reciben entrenamiento adicional para aprender a aplicar los criterios establecidos durante los procedimientos de monitoreo y consistencia que deben realizar.

### **Extensión de las sesiones de codificación**

La codificación de las respuestas abiertas requieren un esfuerzo mental y un alto grado de concentración, lo que no puede ser mantenido por largos periodos, es por ello recomendable que un codificador tenga una jornada de 6 horas por día y que durante este lapso tome dos o tres recesos para tomar café.

Los líderes de mesa deben trabajar un poco más al día para poder desempeñar su labor de monitoreo.

## Logística anterior a la codificación

### Clasificación de los cuadernillos

Luego de revisados y contados contra la lista original de envío, los cuadernillos respondidos deben ser separados de los no utilizados. Luego de lo cual, si no es posible esperar a recibir el total de cuadernillos, se recomienda iniciar la codificación cuando se tenga mínimo la mitad de éstos.

### La selección de los cuadernillos para la codificación múltiple

El Grupo de Asesoría Técnica decidió separar en cada país un número de cuadernillos para ser sometidos al procedimiento de codificación múltiple, el cual varía según la muestra seleccionada y según el dominio al que se ha dado mayor énfasis.

El principio básico en la selección de los cuadernillos para codificación múltiple es que sea de manera aleatoria y que se considere la muestra completa.

El método más sencillo para llevar a cabo la selección de cuadernillos para esta codificación múltiple es usar una proporción basada sobre el número total de cuadernillos esperado.

## Cómo se hace la codificación

Una pequeña línea de números que aparece en la parte superior derecha de cada reactivo en los cuadernillos, corresponde a los posibles códigos que se pueden aplicar a cada respuesta del estudiante en los casos de reactivos de respuesta abierta (esos códigos son explicados en la Guía de Calificación, incluyendo el código “n” que es usado en casos de problemas de impresión).

La codificación es hecha directamente en el cuadernillo (con una excepción que se explica en el apartado de codificación múltiple), circulando el número apropiado del código, de conformidad con lo dispuesto en la Guía de Calificación para la respuesta de cada estudiante en específico.

Los códigos pueden ser:

- 0, 1, 9 y n, los cuales indican incorrecto, correcto, no respondido y no aplicable respectivamente.
- Dentro de los reactivos abiertos existen aquellos que pueden tener respuestas parcialmente correctas, para los cuales se asignó el código 2 para codificar “totalmente correcta” y el 1 para “parcialmente correcta”
- También existen tres grados de respuestas correctas caracterizadas por los códigos 1, 2 y 3.
- Adicionalmente, para discriminar en una respuesta correcta el contenido de la respuesta o el tipo de método de solución usado por el estudiante, se utilizan códigos de doble dígito, tales como 03, 12, 23, 31, 99, etc.

## Números de identificación de los codificadores

A cada codificador debe dársele un número de identificación de conformidad con un formato estándar especificado por el Consorcio, el cual se construye con tres dígitos: el primer dígito representa el dominio del cual es codificador (lectura, matemáticas, ciencias o solución de problemas), los otros dos dígitos son el número consecutivo de codificadores del mismo dominio.

## Diseño de distribución de cuadernillos por codificador

La codificación tiene dos etapas: codificación individual y codificación múltiple, en el primer procedimiento el codificador se encarga de codificar un módulo de todos los cuadernillos que se le asignan. En el caso de la codificación múltiple, los mismos

cuadernillos son codificados por cuatro codificadores distintos, según se describe más adelante.

Los dos procedimientos de codificación son diferentes, pero en ambos casos es complicado. Esta complejidad es con el ánimo de lograr mayor confiabilidad en este proceso.

Primero se debe distribuir los cuadernillos en paquetes que deben ser visiblemente identificados.

Los paquetes pueden ser seleccionados por escuela o por área geográfica y para cada una de éstas, dividirlos por número de cuadernillo (los cuales serán según sean el número de versiones en las que se haya dividido la prueba).

Luego diseñar para la codificación individual, una tabla de distribución mediante la cual cada codificador solamente trabaje una vez cada paquete, es decir, que un codificador no deberá tener en sus manos un mismo paquete dos veces.

Esto implica que cada codificador trabajará un solo módulo cada paquete. El mismo procedimiento se realiza para cada dominio de la prueba.

Un ejemplo gráfico de la distribución de cuadernillos para la codificación individual se muestra en el siguiente cuadro:

Cl	Bklt	Subsets of Schools															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
R1	1,5,7	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16
R2	1,2,6	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
R3	2,3,7	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
R4	1,3,4	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
R5	2,4,5	M13	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
R6	3,5,6	M12	M13	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
R7	4,6,7	M11	M12	M13	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
R8	7,8,9	M10	M11	M12	M13	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8	M9
R9	8,9	M9	M10	M11	M12	M13	M14	M15	M16	M1	M2	M3	M4	M5	M6	M7	M8

Cl = Módulo; Bklt = Cuadernillo; M1, M2, ... = Matemáticas y Número del módulo

## La codificación múltiple

Esta codificación se realiza después de la codificación individual en los cuadernillos que para tal propósito debieron separarse, de manera aleatoria, desde el principio del proceso (tal como se ha indicado). En este caso se procede de la siguiente manera:

- Los cuadernillos se agrupan por su número (según el número de versiones de la prueba y las indicaciones del Consorcio).
- Se ordenan alfabéticamente según el nombre del estudiante.
- Cada grupo de cuadernillos del mismo número se divide en 4 paquetes con igual número de cuadernillos cada uno (si esto es posible por el número de cuadernillos seleccionados para la codificación múltiple).
- En esta etapa la codificación se anota en formatos diseñados para tal efecto las cuales son llamadas cédulas de registro, mismas que contienen los códigos de los reactivos seleccionados por el Consorcio para ser codificados de manera múltiple.
- Cada cédula de registro corresponde a un número de cuadernillo específico y debe ser reproducida tantas veces como número de cuadernillos constituyan cada paquete (de los cuatro en que se dividió cada cuadernillo), luego ese número de cédulas de registro se deberá reproducir por tres, que será el número de veces que se codificará cada cuadernillo en las cédulas.

- Se selecciona a los codificadores que realizarán este procedimiento de codificación, los cuales serán en múltiplos de cuatro de preferencia
- También se nombra a los supervisores quienes además de asesorar a los codificadores, vigilarán el cumplimiento del procedimiento; entregarán a cada codificador las Cédulas de Registro que les corresponda y; recogerán las Cédulas usadas para ordenarlas.
- Cada grupo de cuatro codificará los cuadernillos del mismo número y los cuatro codificarán cada cuadernillo. Es decir que si es un grupo de 100 cuadernillos del número 6, los cuatro codificadores harán su trabajo en esos mismos 100 cuadernillos.
- Para iniciar la codificación, se entrega a cada codificador un paquete del cuadernillo que codificará ese grupo y un paquete de cédulas de registro que corresponde, por nombre y orden, al paquete de cuadernillos que se le ha entregado. Lo mismo sucede para cada uno de los miembros del grupo.
- Procederán a codificar el paquete entregado en las cédulas de registro NO en el cuadernillo y codificando sólo aquellos reactivos que aparecen en la cédula.
- Al terminar, cada codificador entregará al supervisor el paquete de cédulas de registro que utilizó.
- Acto seguido el codificador recibirá del compañero de su grupo que está a su izquierda el siguiente paquete del mismo cuadernillo y el supervisor le entregará el nuevo paquete de cédulas de registro que corresponde al nuevo paquete de cuadernillos. (cada codificador pasará el paquete de cuadernillos que codificó a su compañero situado a la izquierda).
- Estos mismos pasos se repetirán 3 veces

- La cuarta ocasión que reciba un nuevo paquete de su compañero de la izquierda, la codificación se hará DIRECTAMENTE en el cuadernillo, no en las cédulas de registro, codificando además todos los reactivos de los módulos del dominio que contenga dicho paquete de cuadernillos (no sólo los que había marcado en las cédulas).

La elaboración del cuadro de distribución de grupos de codificadores por cuadernillo, se aprecia gráficamente en el ejemplo que aparece en el siguiente cuadro.

Paso	Cuadernillo	Clave de Identificación del Calificador	Módulo para Calif. Múltiple	Módulos que serán Calif. individualmente
1	1	701, 702, 703,	M1,M2	M4
	2	705, 706, 707,	M2,M3	M5
	3	709, 710, 711,	M3,M4	M6,PS1
	4	713, 714, 715,	M4,M5	M7,PS2
2	5	703, 704, 705,	M5,M6	S1,M1
	6	707, 708, 709,	M6,M7	S2,M2
	8	711, 712, 713,	S1,S2	M4
	12	715, 716, 701,	PS1,PS2	M2,S1
3	10	Calificadores	M1, M2	PS2, M6

### Codificación del cuestionario

Este proceso es independiente de la codificación de los cuadernillos de prueba. La codificación requerida para el Cuestionario del Estudiante es para registrar la ocupación del padre y de la madre, así como las expectativas del estudiante. Los códigos para esta tarea son obtenidos del Internacional Standard Classification of Occupations (ISCO).

## Captura de datos, revisión y presentación de archivos

### Captura de datos

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 cuadernillos y los cuestionarios.

El programa tiene integradas fórmulas para revisar la validez de los datos que son ingresados.

Aunque es posible registrar la información en otros paquetes y luego exportarlos al KeyQuest, se recomienda que se haga directamente en este programa para aprovechar muchos de los recursos específicos de PISA.

### Revisión de los datos

El NPM es el responsable de asegurar la calidad de los datos del país antes de que los registros sean presentados al Consorcio.

Las listas con la muestra de escuelas y el formato de seguimiento del estudiante de cada escuela debe ser precisamente cumplimentado e ingresado al KeyQuest.

Cualquier error debe ser corregido antes de ser enviado (copias de los reportes totalmente revisados deben ser presentados junto con los archivos de datos).

### Datos que deben presentarse

Los archivos que deben enviarse al Consorcio son:

- Los datos de los cuadernillos y del cuestionario de contexto
- Datos de los instrumentos internacionales si se usaron
- Los datos de la calificación múltiple
- La lista de la muestra de escuelas
- Las formas de seguimiento del estudiante

## Después de la presentación de los datos

El NPM debe nombrar a un responsable de los datos quien deberá trabajar muy activamente con el centro de procesamiento del Consorcio (en ACER), durante el proceso internacional de revisión y corrección de datos.

Debe además, responder las preguntas que el centro de procesamiento requiera durante este periodo.

## 2.4 Monitoreo de la Calidad

Es esencial que un proyecto con un perfil tan alto y costoso como lo es el proyecto PISA, deba realizarse con altos estándares de calidad.

Esto no sólo requiere que los procedimientos sean cuidadosamente diseñados, sino que además deben ser cuidadosamente monitoreados para asegurar que de hecho se instrumenten a cabalidad.

En caso de que esto no fuese así, es necesario saber en que momento no se dio cumplimiento a lo dispuesto en los procedimientos y en que medida, estar en condiciones de determinar las implicaciones que tendrá dicho incumplimiento sobre la calidad de los datos que se obtengan.

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, pues es él quien instrumenta y guía la operación con lo cual establece el control de calidad.

El programa nacional y las visitas a las escuelas son la base para asegurar la correcta implantación de los procedimientos PISA.

Con el programa de visitas a las escuelas, lo que se logra es prevenir los problemas y asegurar que los datos que se produzcan tengan la calidad requerida para permitir los estudios comparativos entre países.

Existen dos instancias de monitoreo de la calidad en PISA:

- El Centro Nacional de Monitores de Calidad (NCQM), para observar la operación de los procedimientos a nivel nacional.

El Consorcio envía monitores al NPM de cada país antes de los procesos de campo y del estudio principal.

- La Escuela de Monitores de la Calidad (SQMS), son empleados por el Consorcio en los países participantes.

Los SQMS hacen visitas a las escuelas para registrar que tan bien instrumentados están los procedimientos. Durante el procedimiento de campo se visitan pocas escuelas, pero durante el estudio principal las visitas son a un buen número de escuelas.

## Preparación de los procedimientos de monitoreo de la calidad

Miembros del Consorcio realizan visitas antes del procedimiento de campo para diseñar los instrumentos que se requieren para recoger los datos que deberán ser reportados durante el monitoreo.

Con estos propósitos ha sido desarrollada una entrevista estandarizada para ser aplicada al NPM. Este instrumento está basado en aspectos de los procedimientos descritos en el Manual del NPM así como en otros aspectos relevantes, por ejemplo, las relaciones entre las escuelas participantes y el centro nacional.

Se elabora un Manual del Monitor de Calidad, el cual es preparado para el proceso de campo, luego es revisado y aplicado en el estudio principal.

Este Manual incluye:

- Los requerimientos de confidencialidad del monitor
- Formato de registro de datos durante la visita a la escuela
- Agenda para la entrevista con el coordinador de escuela
- Formulario para una entrevista corta con estudiantes cuyo propósito es conseguir sus comentarios sobre los cuestionarios (esta se aplica en el proceso de campo)

## Instrumentación de los procedimientos de monitoreo de calidad

El Consorcio organiza el entrenamiento para el Centro Nacional de Monitores de la Calidad, éstos, a su vez, capacitan a los monitores de la Escuela. Algunos de los aspectos sobre los que se capacita a los monitores son:

- Una visión general sobre el diseño y propósitos del proyecto PISA
- Especial atención sobre las responsabilidades del NPM en la conducción del proceso en sus países.
- Revisión del Manual del Monitor de Calidad

Para la contratación de los monitores (SQMs), se solicita al NPM que presente candidatos al Consorcio, el cual, en su caso, los aprobará y contratará. Las características generales que deberán tener los candidatos a monitores serán las siguientes.

- De preferencia deben estar enterados de los procedimientos y materiales de PISA
- Hablar fluidamente el idioma del país y también el inglés o el francés
- Tener antecedentes en el área educativa o de evaluación
- Pueden ser miembros del Consorcio o individuos específicamente seleccionados para realizar el monitoreo en las escuelas
- Estar en posibilidades de realizar el monitoreo dentro y fuera de su país, (pues en función de los convenios recíprocos, es posible que se requiriera monitorear en otro país)

Una vez contratados y capacitados los (SQMs), sus visitas a las escuelas durante el proceso de campo serán para observar y juzgar:

- La calidad del Manual del Aplicador.
- La claridad de las instrucciones usadas por los aplicadores.
- La necesidad de que algún procedimiento o documentación que será usada en el estudio principal deba ser modificada.

En el estudio principal las visitas se enfocan explícitamente sobre la instrumentación de los procedimientos y el impacto que tiene en los datos el ambiente durante la prueba.

Las visitas de los SQMs a las escuelas en su mayoría son realizadas sin previo aviso y en ellas se observa y registra entre otras cosas:

- Las actividades antes de la administración de la prueba, por ejemplo, la distribución de los cuadernillos.
- La administración de la prueba.
- El registro del aplicador en torno a los acontecimientos en general y el cumplimiento de los procedimientos. Por ejemplo, el registro de coincidencia de la muestra de estudiantes.
- Los comentarios sobre los procedimientos y los materiales.

Una vez que los SQMs completan el reporte de cada escuela visitada, los envían al Consorcio.

## 3. Procesamiento de Datos

### 3.1 Ponderación del estudio y cálculo de la varianza de la muestra

Por el tipo de estudio que representa PISA, el procesamiento de los 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 que 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.

En general, este procesamiento consiste en calcular la ponderación de la población medida y excluida, proporcionar datos que permitan estimar los sesgos del error estándar, llevar a cabo pruebas de significancia, y crear intervalos de confianza apropiados en función del diseño de la muestra de cada país.

La ponderación es necesaria, debido a que los estudiantes incluidos en las muestras del estudio de cada país no son igualmente representativos de la población total de estudiantes, a pesar de haber realizado el muestreo al azar de escuelas y estudiantes.

Hay razones para pensar que las ponderaciones del estudio no son iguales para todos los estudiantes de un país:

- Una muestra de escuelas pudo haber sobremuestreado o submuestreado intencionalmente ciertos sectores de la población total de escuelas; en tales casos se podría analizar por separado, como por ejemplo: una escuela relativamente pequeña, pero políticamente importante para la región; o escuelas geográficamente lejanas, o escuelas con una instrucción o idioma particular.
- La información sobre el tamaño de la escuela en el momento del muestreo puede no ser completamente válida. Si se espera que una escuela sea grande, la probabilidad de selección se basa en la suposición de que una muestra de sus estudiantes podría ser seleccionada. Sin embargo, si la escuela resulta ser demasiado pequeña, entonces todos los alumnos deben ser incluidos y en consecuencia hay una alta probabilidad de ser seleccionados en la muestra, de tal forma que la probabilidad de inclusión en este último caso es más alta que en el primer caso.
- La no respuesta de una escuela cuando no hay reemplazos puede ocurrir, lo cual causaría una sub-representación de esa clase de escuelas, a menos que no se realice un ajuste de la ponderación.
- La no respuesta de un estudiante de las escuelas participantes ocurre en diferentes niveles. Este tipo de estudiantes quedarían sub-representados en los datos, a menos que no se realice la ponderación.
- Ajustar (trimming weights) las ponderaciones para prevenir influencias indebidas de un conjunto relativamente pequeño de muestras de escuelas o estudiantes se necesita si un grupo reducido de estudiantes tuviera mucho mayor peso que el resto de estudiantes del país. Esto causaría estimaciones inestables (grandes errores de muestreo) que no podrían ser estimados correctamente. El ajuste de las ponderaciones introduce un sesgo mínimo en las estimaciones, pero reduce el error estándar.

- Las ponderaciones necesitan ajustarse para que los datos de matemáticas y ciencias reflejen que no todos los estudiantes se evaluaron en cada dominio.

Tomando en cuenta las razones anteriores, PISA realiza procedimientos estadísticos para ponderar los datos y determinar la varianza de la muestra y así asegurar un análisis preciso y confiable de los datos, y deducir inferencias más acordes a las realidades de cada país.

### 3.2 Calibración de los reactivos

Un primer análisis de los datos del estudio de PISA se encamina al análisis y ajuste de las muestras para mejorar las estimaciones y lograr una mayor confiabilidad en las interpretaciones. Otro análisis fundamental, se centra en la validez y confiabilidad de los instrumentos de medición.

El análisis de reactivos se realiza bajo el modelo de respuesta al ítem de Rash.

Este modelo se aplica para:

- Las calibraciones nacionales,
- La calibración internacional,
- La obtención de los puntajes de los estudiantes.

Tanto para las calibraciones nacionales como para las escalas internacionales se usa el modelo de respuesta condicional al ítem, junto con el modelo de población, ya que se asume que los estudiantes fueron muestreados a partir de una distribución normal multivariada.

## Calibraciones nacionales

Estas calibraciones se realizan por separado, es decir país por país, usando datos sin ponderar. Estos resultados son la fuente de información para decidir qué hacer con los reactivos de cada país. Esto significa que un reactivo puede omitirse si presenta características psicométricas pobres en más de ocho países (dodgy item); puede ser que haya reactivos que no se administraron en algunos países, pero funciona bien en la mayoría; o bien un reactivo se comporta bien en cada país, pero muestra interacciones sustanciales entre países, y pareciera que se tratara de reactivos diferentes.

Al revisar las calibraciones nacionales, se pone atención en tres indicadores: a) el ajuste de los reactivos al modelo de escala, b) el poder de discriminación de los reactivos y c) las interacciones reactivo por país.

## Calibración internacional

Para obtener la calibración internacional de los reactivos se aplica el modelo de respuesta condicional al ítem, junto con el modelo de población multivariado, sin usar las variables condicionadas, a una sub-muestra de estudiantes. Esta sub-muestra se obtiene al azar del universo de estudiantes medidos de los países participantes.

## Obtención de los puntajes de los estudiantes

Las competencias (o medidas) de los estudiantes no se observan directamente, más bien son datos que se infieren a partir de las respuestas dadas a los reactivos (respuestas observadas). Existen aproximaciones alternativas para obtener estas inferencias. PISA usa dos aproximaciones:

- La máxima verosimilitud, usando el estimador de verosimilitud ponderado de Warm's (1985), y
- Los valores plausibles.

El estimador de verosimilitud ponderado hace que el puntaje verdadero que un estudiante logra sea más probable. Los valores plausibles son una selección de competencias probables para estudiantes que obtuvieron cada puntaje.

### 3.3 Codificación y estudios de confiabilidad de los calificadores

El estudio de PISA se caracteriza por usar una proporción sustancial de reactivos abiertos en sus pruebas, lo cual obliga a entrenar calificadores o codificadores, cuya función es codificar las respuestas de los estudiantes en apego a las Guías de Calificación diseñadas para este fin. Dada esta circunstancia, PISA establece diversos controles y procedimientos que maximicen la validez y consistencia de la calificación inter e intra país.

PISA establece tres estudios de confiabilidad de la calificación:

- Análisis de la homogeneidad,
- Análisis de los componentes de la varianza, y
- Confiabilidad entre países.

#### Análisis de la homogeneidad

El análisis de la homogeneidad se lleva a cabo con los datos de la prueba piloto y su función es ayudar a los diseñadores de las pruebas a construir escalas confiables y válidas.

Este análisis se enfoca a la prueba piloto, porque muchos de los reactivos se prueban y se calibran para seleccionar los mejores para la aplicación definitiva. En ese sentido, el interés es asegurar que los calificadores logren un nivel de acuerdo razonable en la categorización de las respuestas.

La técnica usada para analizar los datos de la prueba piloto provee una respuesta parcial a los problemas de la categorización y da una indicación del acuerdo entre calificadores para cada reactivo por separado. Cabe resaltar que el conjunto de datos para este análisis se trata como un conjunto de variables nominales o cualitativas.

### **Análisis de los componentes de la varianza**

La prueba piloto concluye con la selección de un conjunto definitivo de reactivos y unas reglas de calificación para cada reactivo. En la aplicación definitiva, el problema es diferente. El principal problema en la aplicación es determinar la varianza total de los puntajes numéricos de la prueba que puede ser atribuida a la variabilidad entre los calificadores.

El conjunto de datos básicos para ser analizados, después de la aplicación consiste en una tabla tridimensional de los puntajes numéricos a la que se le aplica la técnica de análisis de los componentes de la varianza. En suma, esta técnica se realiza con los datos de la aplicación trianual (main study) para examinar los índices de confiabilidad entre países.

La aproximación para estimar la variabilidad en los puntajes en razón de los calificadores es la teoría de la generalizabilidad, cuya explicación se encuentra en Cronbach, Gleser, Nanda and Rajaratnam (1972), Brennan (1992).

## Confiabilidad entre países

El estudio de confiabilidad entre países tiene el propósito de analizar la consistencia en la aplicación de las Guías de Calificación entre países. Mediante esta técnica se investiga la posibilidad de sesgo sistemático en la calificación de las respuestas abiertas.

El objetivo de este estudio es revisar si en general o para algunos reactivos en particular las codificaciones dadas por diferentes grupos nacionales se consideran equivalentes.

Para llevar a cabo esta tarea el Consorcio solicita a los países participantes en la evaluación el envío de una muestra determinada de cuadernillos correspondientes a la calificación múltiple. Estos son recalificados por un grupo entrenado por el Consorcio y sus calificaciones se comparan con las cuatro calificaciones asignadas por los calificadores nacionales a las mismas respuestas de los estudiantes.

Los casos que resultan con una clara discrepancia entre las calificaciones del codificador independiente y los codificadores nacionales son enviados a los investigadores del Consorcio para su análisis.

## 3.4 Procedimientos de limpieza de datos

La limpieza de los datos es un componente esencial del Programa de Control y Aseguramiento de la Calidad de PISA. Es de suma importancia que el Consorcio identifique las anomalías e inconsistencias en la base de datos enviada por cada país y que ningún error se introduzca durante las fases de limpieza y análisis.

Después de haber calificado y capturado toda la información de la población evaluada, el Consorcio solicita a los Administradores Nacionales del Proyecto enviar sus datos en un

software específico, KeyQuest. Los datos remitidos corresponden a los cuadernillos, los cuestionarios, las cédulas de la calificación múltiple, la lista de escuelas y las cédulas de registro de alumnos.

Los datos se verifican en diferentes puntos (verificación de integridad) desde que son capturados en el KeyQuest, puesto que este software contiene reglas de validación (verificación de rangos) especificadas para cada variable, y una variable de dato (variable datum) que sólo es aceptada si se satisfacen las reglas de validación. Para prevenir registro duplicados, el software referido cuenta con unas llaves primarias (primary keys) que son asignadas a variables como: el estrato y las identificaciones de la escuela y del estudiante.

Antes de empezar a capturar datos de los cuadernillos y de los cuestionarios, se deben ingresar los datos al módulo de las cédulas de registro de alumnos. Este módulo contiene la identificación completa del estudiante, tal y como debe aparecer en el cuadernillo que el estudiante debe responder durante la evaluación.

Una vez que el proceso de captura se completa, se le solicita al Administrador Nacional del Proyecto aplicar algunos procedimientos de verificación, usando el KeyQuest, antes de enviar la base de datos al Consorcio, a fin de rectificar cualquier error. Estos errores pueden ser inconsistencias entre la lista de escuelas y los cuestionarios; las cédulas de registro de alumnos y los cuadernillos; las cédulas de registro de alumnos y los cuestionarios del estudiante; los cuadernillos y los cuestionarios del estudiante; y los datos de confiabilidad de la calificación múltiple conforme al diseño internacional para detectar si se calificaron más de cuatro veces los estudiantes por cuadernillo.

## 4. Construcción de Escalas

### 4.1 Construcción de escalas de competencias

PISA busca reportar resultados de competencias que estén basadas en una teoría científica, busca además que dichos resultados sean interpretables en términos de políticas.

Para ello desarrolla escalas que reflejan, por niveles, los resultados de los estudiantes evaluados.

Los niveles en estas escalas requieren satisfacer dos consideraciones en su desarrollo:

- PISA debe entregar resultados individuales para cada país y por cada uno de los dominios cognoscitivos evaluados. Esto significa que las escalas múltiples deben ser útiles para esos propósitos específicos.
- La descripción de las competencias debe arrojar luz sobre las tendencias a través del tiempo, a pesar de que la cantidad de información disponible para soportar la descripción detallada de la variación de las capacidades, dependa de si un dominio en particular es mayor o menor para un ciclo de medición en particular (debido al énfasis que se da a un dominio en cada ciclo de evaluación).

Las decisiones acerca del desarrollo de las escalas necesitan reconocer esta variación y deben facilitar la descripción de cualquier cambio en los niveles de competencias logradas por los países de un ciclo de medición al próximo.

El desarrollo de un método para describir las competencias de cada dominio seleccionado por PISA, tomó más de un año en estar listo para la emisión de un reporte de resultados.

Para desarrollar los paquetes de escalas de competencias por dominio de la prueba, Grupos Funcionales de Expertos (FEGs) trabajaron con el Consorcio, el cual consultó al Comité de Países Participantes (BPC), a los Administradores Nacionales del Proyecto (NPMs) y al Grupo Técnico Consultivo (TAG) de PISA. Las escalas fueron presentadas a los (BPCs) para su aprobación, en abril de 2000.

## Desarrollo de la descripción de las escalas

El desarrollo de la descripción de las escalas de competencias PISA, debió pasar por un proceso que implicó una serie de fases:

### *Fase 1:* Identificación de posibles subescalas

Requirió la reunión de expertos en cada dominio para articular posibles escalas que pudieran ser reportadas.

Fueron consideradas varias escalas para cada dominio, las cuales claramente surgían del marco de referencia correspondiente, por lo que fueron vistas como potencialmente significativas para los propósitos de reporte y retroalimentación, sin embargo necesitaban ser respaldadas respecto a sus propiedades de medición.

Por ejemplo para lectura se consideraron dos opciones. Estas escalas estaban basadas en dos posibles agrupamientos de cinco aspectos de lectura: extracción de información; adquisición de un claro entendimiento; desarrollo de una interpretación; reflexión sobre el contenido y; reflexión sobre la forma del texto. (Estrictamente hablando estas son subescalas, pero para simplificar las llamaremos escalas).

### **Fase 2:** Asignación de reactivos a las escalas

Esta fase consistió en asociar cada reactivo del estudio principal con cada escala considerada. Esto fue hecho con la participación de especialistas de los FEGs en cada materia del dominio, los diseñadores de la prueba y el equipo del Consorcio. También en ocasiones se contó con la participación los NPMs que estaban involucrados.

Los análisis estadísticos de los puntajes de los reactivos en el estudio de campo, fueron también útiles para identificar el grado con el cual éstos se podrían ubicar mejor en cada subescala dentro de la escala y validar el trabajo de los expertos del dominio.

### **Fase 3:** Auditoría de habilidades

Esta fase fue de análisis detallado de los expertos para cada reactivo (en el caso de reactivos con crédito parcial, para cada puntaje alcanzado dentro de dicho reactivo), en relación con la relevancia definida en el marco de referencia de cada dominio.

Las habilidades y los conocimientos requeridos para alcanzar cada puntaje fueron identificados y descritos

Esta etapa implicó discusión y negociación entre los involucrados, la realización de borradores y el progresivo refinamiento de éstos, con base en las aportaciones y la retroalimentación de los expertos.

### **Fase 4:** Análisis de los datos del proceso de campo

Para cada grupo de escalas considerado, los datos obtenidos del proceso de campo de cada reactivo (que luego fue seleccionado para el estudio principal) fue analizado usando la técnica de la respuesta al reactivo con el fin de obtener la dificultad aproximada para cada umbral de logro, por cada reactivo en cada subescala.





**Fase 5:** Definición de las dimensiones

La información emanada del análisis del dominio específico hecho por los expertos (Fase 3) y el análisis estadístico (Fase 4), fueron combinados.

Para cada grupo de escalas consideradas, el puntaje de cada reactivo fue ordenado de acuerdo con el tamaño de sus umbrales asociados y luego relacionados con las descripciones de los conocimientos y habilidades con que están asociados, asignando una jerarquización de conocimiento y habilidades según la definición de la dimensión.

Con este procedimiento las habilidades se agrupan naturalmente, lo cual permite una base para entender cada dimensión y describir los niveles de competencias.

### **Fase 6:** Refinación de los datos del estudio principal

Una vez que los datos del estudio principal están disponibles, la información obtenida del análisis estadístico en torno a la dificultad relativa de los umbrales de los reactivos es actualizada.

Esto hace posible que los grupos de trabajo (FEGs y otros) puedan hacer un resumen y revisión de la fase 5.

Las descripciones preliminares y los niveles asignados, son entonces revisados bajo la óptica que proporciona la información técnica adicional (obtenida de los datos del proceso de campo), la cual es proveída por el Grupo de Asesores Técnicos.

### **Fase 7:** Validación

Son dos formas de validación las que fueron consideradas y usadas por los tres grupos de trabajo.

Un método fue aportado por los expertos conocedores (por ejemplo maestros o miembros de grupos expertos en la materia), con material que les permitía juzgar los reactivos PISA contra los niveles descritos, o contra los indicadores que soportan la descripción de dichos niveles.

El segundo método consistió en un proceso de consulta en el que involucró a todos los países que participan en el PISA, a través de sus respectivos NPM. Esta aproximación para la validación es útil en tanto los usuarios de las descripciones las encuentren informativas.

## **Definición de los niveles de competencia**

El desarrollo de la descripción de niveles se desenvuelve en dos etapas. La primera se realiza después de la descripción de las escalas. Se basa en el análisis sustantivo de los reactivos PISA en relación con los aspectos cognoscitivos que soporta cada dominio.

Esto produce niveles de competencia que reflejan las observaciones del desempeño de estudiantes y un análisis detallado de las exigencias cognoscitivas de los reactivos de PISA.

La segunda etapa implica decisiones en torno a la ubicación de los puntos de corte para los niveles y como asociar estudiantes con cada nivel. Esta etapa es a la vez técnica y muy práctica, pues permite interpretar lo que significa “estar en un nivel” y tienen significativas consecuencias en los reportes de resultados nacionales e internacionales.

Fueron considerados varios principios para desarrollar y establecer un significado útil al hecho de “estar en un nivel” y por lo tanto, para determinar una fórmula para la localización de los puntos de corte entre niveles y el correspondiente apareamiento de estudiantes a éstos:

- El significado común de niveles debe ser desarrollado y elevado de rango. Es importante que primero se considere que los conocimientos que el estudiante posee en las habilidades que mide PISA, son un continuo: No hay puntos de ruptura naturales que marquen líneas divisorias entre etapas a lo largo de ese continuo. La división de ese continuo en niveles es un proceso arbitrario aunque conveniente para propósitos de evaluación.

Es útil definir fases o niveles a lo largo del continuo, porque esto nos permite comunicar las competencias de los estudiantes en términos que no sean sólo numéricos.

- Por ejemplo, los estudiantes ubicados hasta el final de un nivel, podían completar al menos 50% de las tareas correctamente en una prueba que considerara dicho nivel, mientras que los estudiantes a la mitad o arriba de cada nivel podría esperarse que alcanzaran una mayor tasa de éxito. En el límite superior de un nivel podrían estar los “maestros” de ese nivel.

Pero aunque un estudiante esté hasta el borde superior de un nivel, también estaría en el límite inferior del siguiente nivel superior, donde, de acuerdo con el razonamiento inicial, podría resolver con toda seguridad por lo menos el 50% de cualquier tarea definida en ese nivel más elevado.

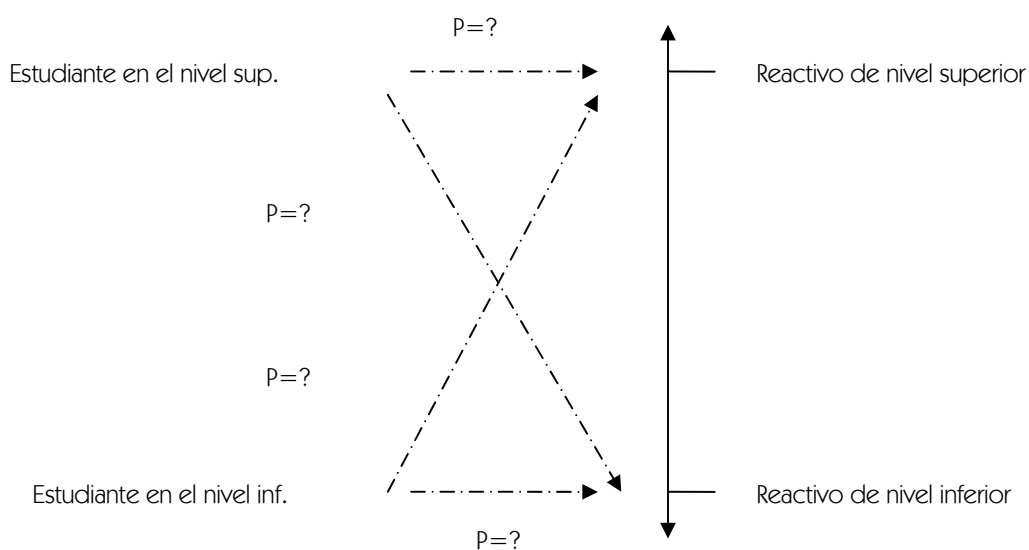
- Adicionalmente, el significado de “estar en un nivel” en una escala dada, puede ser más o menos consistente para cada nivel. Esto es, si se consideran en toda su extensión las definiciones y descripciones sustantivamente basadas de los niveles, los puntos de corte pueden crear niveles con una longitud más o menos constante. Claro que esto no aplica para los niveles de competencias más bajos y los más altos.
- Con una aproximación más o menos consistente para definir los niveles de las diferentes escalas, sus amplitudes pueden no ser exactamente las mismas para cada escala de competencias, sin embargo si pueden tener el mismo tipo de interpretación.

Una manera de instrumentar estos principios ha sido desarrollada. Este método relaciona las dos variables mencionadas en los párrafos anteriores con una tercera variable.

Estas tres variables pueden ser las siguientes:

- El éxito esperado de un estudiante en un nivel particular de una prueba que contiene reactivos de ese nivel.
- El ancho de los niveles en la escala, determinados ampliamente por las consideraciones sustantivas de las demandas cognitivas de los reactivos de cada nivel y las observaciones del desempeño del estudiante en los reactivos y;
- La probabilidad de que un estudiante en la mitad de un nivel pueda contestar correctamente los reactivos de dificultad promedio para dicho nivel, algunas veces referido como valor RP para la escala (donde RP significa Respuesta Probable)

El cuadro siguiente resume la relación entre estas tres variables matemáticas. Muestra una línea vertical que representa una parte de la escala que ha sido definida, uno de los niveles delimitados en la escala, un estudiante en los dos extremos del nivel y, la referencia a un reactivo hasta el límite superior y otro en el inferior. Las líneas punteadas relacionan a los estudiantes con los reactivos que han sido calificados  $P = ?$ , para indicar la probabilidad asociada con ese estudiante que respondió correctamente ese reactivo.



### Ejemplo de escala para alfabetización en ciencias (dominio ciencias)

Descripción del nivel más alto

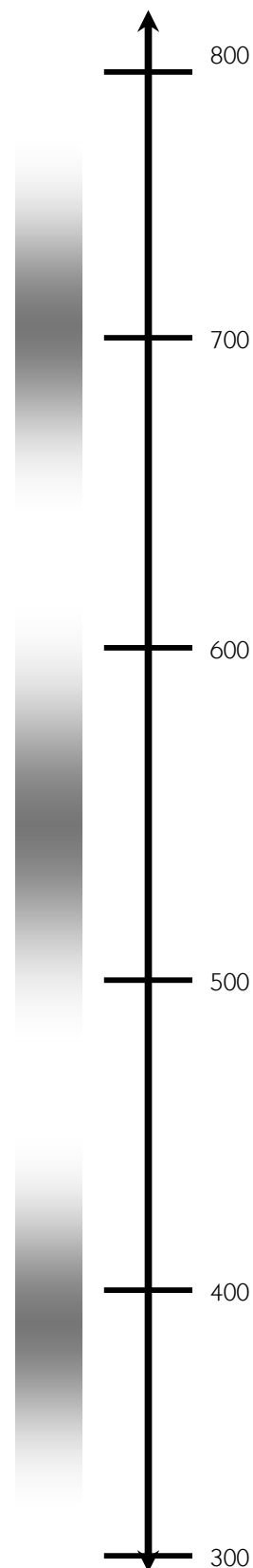
- Crear o usar modelos conceptuales para hacer predicciones o dar explicaciones
- Analizar investigaciones científicas en relación con, por ejemplo, diseños experimentales, identificación de las ideas que están siendo probadas
- Relacionar datos como evidencia para evaluar puntos de vista alternativos o diferentes perspectivas
- Expresar argumentos científicos y descripciones detalladas y precisas

Descripción del nivel medio

- Usar conceptos científicos al hacer predicciones o dar explicaciones
- Reconocer interrogantes que pueden ser respondidas con investigaciones científicas e identificar detalles que tienen implicaciones científicas
- Seleccionar información relevante para construir razonamientos en serie para sacar o evaluar conclusiones

Descripción del nivel más bajo

- Recordar simples conocimientos basados en hechos (por ejem. Nombres, hechos, terminología, reglas simples)
- Usar conocimientos comunes para construir o evaluar conclusiones



## 4.2 Base de datos internacional

### Archivos en la base de datos

La base de datos internacionales de PISA está constituida por cinco archivos: cuatro de ellos a nivel de estudiantes y uno a nivel de escuelas.

### Archivos de estudiantes

Los archivos incluyen los datos de los cuestionarios de contexto y del desempeño de cada estudiante en los dominios evaluados.

El archivo de estudiantes tiene disponible la siguiente información de cada estudiante:

- Identificación del país, la escuela y el estudiante.
- Las respuestas del estudiante en el cuestionario de contexto.
- Los índices derivados de las respuestas del estudiante en el cuestionario.
- Los puntajes alcanzados por el estudiante en cada dominio evaluado.
- La ponderación del estudiante y el factor de ajuste aplicado al país para los pesajes de cada dominio.

### El archivo de escuela

Existe la siguiente información para cada escuela que participó en la evaluación:

- Las variables de identificación de la escuela.
- Las respuestas emitidas en el cuestionario de la escuela.
- Los índices que se derivan de las preguntas originales en el cuestionario de la escuela.

- El pesaje de la escuela.

La base de datos de la escuela ha sido ajustada para que no aparezcan las escuelas que no respondieron y es proporcionada al final del archivo de escuela.

## El archivo de datos de los reactivos de la evaluación

El archivo de conocimientos tiene todos los reactivos incluidos en la prueba y muestra las respuestas de los estudiantes tanto en el formato de un dígito como en el de dos.

Las respuestas a los reactivos de respuesta abierta pueden proporcionar valiosa información acerca del pensamiento e ideas de los estudiantes, lo cual retroalimenta la planeación curricular.

## Información incluida en la base de datos

A nivel estudiante

- Todos los estudiantes que estuvieron en una de las dos sesiones de evaluación (para evaluación de conocimiento o para cuestionario de contexto)
- En el caso de estudiantes que solamente asistieron a la sesión de cuestionario de contexto, se incluyen sólo aquellos que respondieron las preguntas sobre las ocupaciones de la madre y el padre.

A nivel escuela

- Todas las escuelas participantes que por lo menos hayan evaluado el 25% de la muestra de estudiantes elegida, a pesar de que la escuela devuelva el cuestionario.

## Presentación de los datos omitidos

La codificación de los datos distingue entre cuatro diferentes tipos de datos omitidos:

- Reactivo no respondido: cuando no responde una pregunta el estudiante o el director de escuela. El código puede ser 9, 99 o 999, según sea el número de dígitos indicado en el reactivo.
- Respuestas múltiples o inválidas: esta posibilidad se presenta en los reactivos de opción múltiple. Los códigos posibles son 8, 98, 998 y así sucesivamente según sea el número de dígitos del reactivo.
- No aplicable: Es aplicable cuando no es posible que el estudiante conteste el reactivo por problemas de impresión del cuadernillo o por que éste ha sido mutilado. En el cuadernillo estos casos son codificados con la letra “n” y el código es 7, 97, 997, según sea el número de dígitos para el reactivo.
- Reactivos que no alcanzó a contestar: Son todos aquellos que consecutivamente no fueron contestados empezando por el último del cuadernillo y se les asigna el código “r”, exceptuando el primer reactivo no contestado de la serie, al cual se le asignará código “n”

## Identificación para estudiantes y escuelas

La identificación del estudiante está compuesta por tres variables:

- El país, el cual es codificado por PISA según el código de países ISO 3166
- La identificación de la escuela el (SCHOOLID)
- La identificación del estudiante etiquetada como (STIDSTD)
- En una cuarta variable se incluyó la identificación por estado de la república

## Los archivos del estudiante

Dos tipos de índices se obtienen de los archivos del cuestionario del estudiante. El primer grupo se basa en la transformación de una variable o sobre una combinación de información incluida en dos o más variables.

El segundo grupo es el resultado de una escala de Rasch y consiste en ponderar el porcentaje de estimado de índices: 15 índices del cuestionario del estudiante; 14 índices de la opción internacional de cuestionario de cruzamiento de currículo y; 3 índices del cuestionario de información tecnológica.

Estos dos tipos de estimaciones de realizaciones de los estudiantes son incluidas para cada dominio cognoscitivo y por cada subclasificación (por ejemplo: extracción de información, interpretación de textos y reflexión y evaluación, en el caso de lectura).

Es recomendable que este grupo de valores sean usados cuando se analice y reporten estadísticas a nivel de la población.

## Los archivos de conocimiento.

El archivo con la información de la prueba contiene las respuestas individuales de los estudiantes de todos los reactivos usados para la calibración y en la generación de valores verosímiles. Todos los reactivos respondidos en este archivo tienen un formato de un dígito.

En este archivo, los reactivos son clasificados por dominio y clasificados también alfabéticamente dentro de cada dominio.

# 4

## PISA 2003 en México

En este capítulo se describen las etapas y actividades realizadas para la aplicación y la calificación de las pruebas de PISA aplicadas en México, durante mayo-junio del 2003. Está dividido en tres partes: la aplicación, la codificación y la digitación.

Desde la primera aplicación de PISA, México ha estado involucrado en la implementación de las pruebas piloto y las aplicaciones. Hasta ahora México ha sido uno de más de los países que se suman al proyecto PISA, con el convencimiento de que someterse a mediciones de esta envergadura son cruciales 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.

## 1.1 Aplicación

### Capacitación

El 10 y 11 de abril 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, las 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 en cuanto a:

- El periodo de la aplicación,
- La entrega de las listas de alumnos seleccionados conforme a la muestra de escuelas,
- La entrega del calendario de la aplicación,
- El 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 las siguientes:

- De abril a mayo se tradujeron al español los cuadernillos de conocimientos. Se revisaron y ajustaron los cuestionarios de contexto y los manuales del aplicador y del coordinador escolar, que ya habían sido traducidos al español.

- En el periodo de finales de abril y principios de 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 entonces integrar 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 35,730 alumnos de 1,170 escuelas.

### **Población evaluada (alumnos y escuelas)**

Las pruebas PISA se aplicaron a un **total de 30,087 alumnos** de 15 años de edad, **de 1,131 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. La distribución de escuelas y de alumnos a evaluar y evaluados se muestra en el **Anexo 1**.

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

La tasa de participación de alumnos fue de 84%, en tanto que la de escuelas fue de 97% (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% y para alumnos de 80%. La comparación entre las tasas de participación obtenidas y requeridas se muestra en el siguiente cuadro.

Cuadro 1. 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%.

Las entidades con una tasa de participación de alumnos inferior al 80% 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 las cajas con los 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 los cuestionarios para la escuela, y se separó el material utilizado y no utilizado.

## 1.2 Proceso de Codificación

### Traducción y preparación de Guías y Manuales

A partir de la primera quincena de junio y hasta julio se tradujeron las cuatro Guías de Calificación y los cuatro Manuales de Capacitación, correspondientes a Matemáticas, Solución de Problemas, Ciencias y Lectura. La traducción de estos materiales estuvo lista en la primera quincena de julio.

Las traducciones fueron enviadas a PISA para su revisión y aprobación en la última semana de agosto. Las observaciones remitidas por PISA (**Ver Anexo 2**) se recibieron en la siguiente semana y fueron incorporadas a las Guías y los Manuales.

## Planeación y organización de la codificación

A partir de julio se inició la planeación del proceso de la codificación. Se llevaron a cabo varias reuniones de staff con el propósito de:

- analizar y definir el ordenamiento necesario para realizar la codificación individual y múltiple,
- definir si la calificación iba a realizarse de manera simultánea en todas las áreas de evaluación o iniciar con lectura,
- solicitar la prórroga para la entrega de las base de datos a PISA,
- proponer el cambio del NPM y acceder a las páginas WEB confidenciales,
- definir la cantidad y el perfil de los codificadores, y la forma de reclutamiento,
- establecer la organización y las necesidades de recursos materiales y físicos,
- determinar el espacio físico necesario para concentrar y manejar los cuadernillos, así como para tener a los codificadores trabajando en grupos,

## Asesoría externa

Una primera acción efectuada fue solicitar la asesoría presencial de un representante de PISA. Se logró la visita del Dr. Wolfran Schulz de ACER, quien estuvo con el staff el 25 de julio del 2003. La sesión de trabajo duró ocho horas en la que se analizaron las dudas del grupo de trabajo y se le presentaron los avances en relación con la planeación de la calificación. El resultado de esta asesoría fue la aclaración de dudas en cuanto al proceso y algunas preguntas que quedaron sin resolver en ese momento fueron resueltas vía correo electrónico, después de ser consultadas en la instancia respectiva.

## 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 20 personas: 16 calificadores y cuatro supervisores; y para matemáticas 45 personas: 40 calificadores y cinco supervisores,
- Integrar para lectura cuatro grupos de cuatro personas cada uno y para matemáticas 10 grupos de cuatro personas,
- Seleccionar personas de reserva para que en caso de que alguno de los contratados no asistiera, se pudieran 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 (**Ver Anexo 4**). 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 entrevista.

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.

### **Organización de los materiales de la aplicación**

Durante tres semanas, los cuadernillos fueron organizados en cajas por entidad y versión de cuadernillo, identificando número de cuadernillos por caja. Estas cajas se dispusieron en anaqueles para ser manejados fácilmente durante la codificación. Los cuestionarios de la escuela y del estudiante se colocaron en cajas y puestos en un área diferente a la de los cuadernillos. Además, se separaron los 900 cuadernillos para la calificación múltiple.

### **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, 100 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.

### **Diseño de controles para la codificación**

Se diseñaron tres tipos de controles en papel: para el supervisor, para la persona encargada de entregar y recibir los cuadernillos; y para el registro de asistencia del personal encargado de la codificación y supervisión.

El control para la supervisión consistió en una bitácora diseñada para el control de la calidad de la codificación, en éste se registraban, por día, las dudas de los codificadores

en cuanto a la asignación de los códigos, además de observaciones referentes a problemas con los cuadernillos (errores de impresión, paginado, etc.) u otro tipo de observaciones.

Para la entrega y recepción de los cuadernillos, se diseñó un control que contenía la distribución de las versiones de los cuadernillos por Entidad para cada calificador, de acuerdo con el ordenamiento establecido para lectura y matemáticas, solución de problemas y ciencias. Además, se elaboró otro control por codificador para que se anotara la cantidad de cuadernillos recibidos y entregados, la firma del codificador, el avance logrado y si fuera el caso, las preguntas pendientes.

Otro control fue el de asistencia, en éste se registraba la hora de entrada y salida, y la firma del codificador. Este control era por codificador.

## 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 los cuatro restantes para la propia codificación múltiple. El proceso de codificación múltiple 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 era de 9 a 15:30 hrs. con un receso de 15 minutos a la mitad de la misma.

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 el 13 y 14 de octubre.

### 1.3 Digitación

A finales de septiembre se inició la digitación de los cuestionarios de contexto aplicados. Se digitaron alrededor de 100,000 documentos entre los cuestionarios, las cédulas de registro de alumnos, los cuadernillos, reporte de aplicación y la lista de programas de estudio. Esto implica más de 20 millones de golpes de máquina a digitar.

Para el proyecto PISA, 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 de México, la evaluación estandarizada y además conformar grupos de trabajo dedicados al diseño de reactivos y 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 que se implanten en la educación básica.

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

## Referencias

OECD Programme for International Student Assessment (2001). *Knowledge and Skills for Life. First Results From Pisa 2000*. Paris, OECD.

\_\_\_\_\_ *Reading for Change. Performance and Engagement Across Countries. Results from Pisa 2000*. Paris, France.

\_\_\_\_\_ (2002). Muestra de reactivos empleados en la evaluación PISA 2000. Aptitudes para la lectura, matemáticas y ciencias. Aula XXI. Santillana. México.

\_\_\_\_\_ (2002). Conocimientos y Habilidades para la Vida. Primeros Resultados de Pisa 2000. Traducción realizada por la Dirección General de Evaluación para consulta interna, del resumen disponible en: [www.pisa.oecd.or](http://www.pisa.oecd.or)

Martínez Rizo, Felipe (2003). *Los Resultados de las Pruebas Pisa. Elementos para su interpretación*. Ponencia presentada en el “Curso de Actualización para Periodistas” coordinado por el Instituto Nacional para la Evaluación de la Educación, la Organización de Estados Iberoamericanos y el Observatorio Ciudadano de la Educación, en la Ciudad de México el día 3 de Julio de 2003.

## Siglas

Las siguientes siglas son usadas en este documento:

SIGLAS	REFERENCIA EN INGLÉS	REFERENCIA EN ESPAÑOL
ACER	Australian Council for Educational Research	Consejo Australiano para la Investigación Educativa
BPC	PISA Board of Participating Countries	Consejo de Países Participantes de PISA
DIF	Differential Item Functioning	
ENR	Enrolment of 15-year-olds	Inscripción de Estudiantes de 15 años
ETS	Educational Testing Service	Servicio de Evaluación Educativa
FEG	Functional Expert Group	Grupo Funcional de Expertos
I	Sampling Interval	Intervalo de la Muestra
IALS	International Adult Literacy Survey	Evaluación Internacional de la Alfabetización Adulta
IEA	International Association for the Evaluation of Educational Achievement	Asociación Internacional para la Evaluación de los Logros Educativos
ISCED	International Standard Classification of Education	Estándar Internacional de Clasificación de la Educación
ISCO	International Standard Classification of Occupations	Estándar Internacional de Clasificación de Ocupaciones
IT	PISA questionnaire on computer familiarity	Cuestionario PISA de Familiaridad con la Computación
NCQM	National Centre Quality Monitor	Centro Nacional de Monitoreo de la Calidad
NPM	National Project Manager	Administrador Nacional del Proyecto
PISA	Programme for International Student Assessment	Programa para la Evaluación Internacional de Estudiantes
RN	Random Number	Número Aleatorio
SC	School Co-ordinator	Coordinador de Escuela
SD	Standard Deviation	Desviación Stándar
SEM	Structural Equation Modelling	
TA	Test Administrator	Aplicador
TAG	Technical Advisory Group	Grupo Técnico Asesor
TIMSS	Third International Mathematics and Science Study	Tercer Estudio Internacional de Matemáticas y Ciencias
TIMSS-R	Third International Mathematics and Science Study - Repeat	Repetición del Tercer Estudio Internacional de Matemáticas y Ciencias

# INEE

Instituto Nacional de Evaluación  
Para la Educación

Bartolache 1862, Esq. José María Rico  
México D. F.  
C. P. 03100  
Tel. (52) 55. 5524.1840

[www.inee.com.mx](http://www.inee.com.mx)