

Evaluaciones nacionales del rendimiento académico

VOLUMEN 2

Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico

Prue Anderson
George Morgan



GRUPO BANCO MUNDIAL

**Desarrollo de pruebas
y cuestionarios para
una evaluación
nacional del
rendimiento
académico**

Evaluaciones nacionales del rendimiento académico

VOLUMEN 2

Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico

Prue Anderson y George Morgan

Vincent Greaney y

Thomas Kellaghan, editores de la serie



GRUPO BANCO MUNDIAL

© 2016 Banco Internacional de Reconstrucción y Fomento/Banco Mundial
1818 H Street NW, Washington, DC 20433
Teléfono: 202-473-1000; Internet: www.worldbank.org

Algunos derechos reservados
1 2 3 4 19 18 17 16

La presente obra fue publicada originalmente por el Banco Mundial en inglés en 2008, con el título *Developing Tests and Questionnaires for a National Assessment of Educational Achievement*. Vol. 2 of *National Assessments of Educational Achievement*. En caso de discrepancias, prevalecerá el idioma original.

El presente documento ha sido realizado por el personal del Banco Mundial, con aportaciones externas. Las opiniones, las interpretaciones y las conclusiones aquí expresadas no son necesariamente reflejo de la opinión del Banco Mundial, de su Directorio Ejecutivo ni de los países representados por este. El Banco Mundial no garantiza la exactitud de los datos que figuran en esta publicación. Las fronteras, los colores, las denominaciones y demás datos que aparecen en los mapas de este documento no implican juicio alguno, por parte del Banco Mundial, sobre la condición jurídica de ninguno de los territorios, ni la aprobación o aceptación de tales fronteras.

Nada de lo aquí contenido constituirá ni podrá considerarse una limitación ni una renuncia de los privilegios y las inmunidades del Banco Mundial, todos los cuales están reservados específicamente.

Derechos y autorizaciones



Esta publicación está disponible bajo la licencia Creative Commons Reconocimiento 3.0 IGO (CC BY 3.0 IGO): <http://creativecommons.org/licenses/by/3.0/igo>. La licencia Creative Commons Reconocimiento permite copiar, distribuir, comunicar y adaptar la presente obra, incluso para fines comerciales, con las siguientes condiciones:

Cita de la fuente. La obra debe citarse de la siguiente manera: Anderson, Prue, y George Morgan. 2016. *Evaluaciones nacionales del rendimiento académico*. Volumen 2: *Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico*, Vincent Greaney y Thomas Kellaghan, editores. Washington, DC: Banco Mundial. DOI:10.1596/978-1-4648-0745-9. Licencia: Creative Commons Reconocimiento CC BY 3.0 IGO.

Traducciones. En caso de traducirse la presente obra, la cita de la fuente deberá ir acompañada de la siguiente nota de exención de responsabilidad: “La presente traducción no es obra del Banco Mundial y no deberá considerarse traducción oficial de este. El Banco Mundial no responderá por el contenido ni los errores de la traducción”.

Adaptaciones. En caso de que se haga una adaptación de la presente publicación, la cita de la fuente deberá ir acompañada de la siguiente nota de exención de responsabilidad: “Esta es una adaptación de un documento original del Banco Mundial. Las opiniones y los puntos de vista expresados en esta adaptación son exclusiva responsabilidad de su autor o de sus autores y no son avalados por el Banco Mundial”.

Contenido de terceros. Téngase presente que el Banco Mundial no necesariamente es propietario de todos los componentes de la obra, por lo que no garantiza que el uso de dichos componentes o de las partes del documento que son propiedad de terceros no violará los derechos de estos. El riesgo de reclamación derivado de dicha violación correrá por exclusiva cuenta del usuario. Si se desea reutilizar algún componente de esta obra, es responsabilidad del usuario determinar si debe solicitar autorización y obtener dicho permiso del propietario de los derechos de autor. Como ejemplos de componentes se puede mencionar los cuadros, los gráficos y las imágenes, entre otros.

Toda consulta sobre derechos y licencias deberá enviarse a la siguiente dirección: Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; correo electrónico: pubrights@worldbank.org.

ISBN (edición impresa): 978-1-4648-0745-9
ISBN (edición electrónica): 978-1-4648-0746-6; 978-0-8213-7497-9 (inglés)
DOI: 10.1596/978-1-4648-0745-9

Diseño de la portada: Naylor Design, Washington DC



ÍNDICE

PRÓLOGO	xi
ACERCA DE LOS AUTORES Y EDITORES	xv
AGRADECIMIENTOS	xvii
SIGLAS	xix

Parte 1 **Elaboración de las pruebas**

1. INTRODUCCIÓN	3
2. DESARROLLO DE UN MARCO DE EVALUACIÓN	11
Documento de especificaciones técnicas de la prueba o tabla de especificaciones	12
Validez	19
Idioma de la prueba	19
Formato del ítem	20
Población escolar a evaluar	27
Informe de resultados	28
Contextos	29
Notas	30

3. REDACCIÓN DE LOS ÍTEMS	31
Dificultad de los ítems	33
Sesgo de los ítems	34
Material de estímulo	35
Formato de los ítems	38
Ítems de práctica	51
Diagramación y diseño de los ítems	52
El equipo de redacción de ítems	58
Comités encargados de revisar los ítems	62
Otros revisores	64
Seguimiento de los ítems	65
Notas	68
4. ENSAYO DE LOS ÍTEMS	69
Diseño del formulario de la prueba previa	73
Impresión y revisión de la prueba previa	77
Implementación de la prueba previa	80
Corrección de la prueba previa	81
Confiabilidad	87
5. SELECCIÓN DE LOS ÍTEMS DE LA PRUEBA	89
Nota	94
6. PRODUCCIÓN DE LA PRUEBA DEFINITIVA	95
Diseño de la prueba definitiva	95
Impresión y corrección	99
7. CALIFICACIÓN MANUAL DE LOS ÍTEMS	103

Parte 2

Elaboración de cuestionarios

8. DISEÑO DE CUESTIONARIOS	111
Contenido del cuestionario	113
Especificaciones técnicas del cuestionario	117
Ítems de los cuestionarios	119
Formato de los ítems	121
Idioma del cuestionario	122
Encuestados	122
Administración del cuestionario	123
Plan de análisis de los datos	123

9. REDACCIÓN DE ÍTEMS PARA CUESTIONARIOS	125
Preguntas	126
Enunciados	126
Categorías de respuesta	127
Tratamiento de temas delicados	129
Diseño del cuestionario	129
Revisión de los cuestionarios	130
10. CODIFICACIÓN DE LAS RESPUESTAS DEL CUESTIONARIO	133
Preparación de los cuestionarios para la captura de datos	135
Respuestas ambiguas o falta de código	135
11. COTEJO DE LOS CUESTIONARIOS CON LOS DATOS DE LA PRUEBA	137
Cuestionarios para estudiantes	137
Cuestionarios para padres	138
Cuestionarios para docentes y directores	139
Parte 3	
Diseño de un manual para la administración de la prueba	
12. MANUAL DEL EXAMINADOR	143
Contenidos del manual	144
Uso del manual	145
Características del manual	146
¿Cuánta información se necesita?	148
Preguntas de práctica	149
Ensayo	150
Revisión	151
13. EL EXAMINADOR	153
Elección del examinador	153
Cumplimiento de las instrucciones	154
Control de calidad	156
Lista de verificación del examinador	157
14. INFORMACIÓN A LAS ESCUELAS SOBRE LA EVALUACIÓN NACIONAL	161

APÉNDICES	163
A. GLOSARIO	163
B. LECTURAS ADICIONALES	171
C. EJEMPLOS DE ÍTEMS DE PRUEBAS, ÍTEMS DE CUESTIONARIOS Y MANUALES DE ADMINISTRACIÓN EN EL CD	175
Ítems de pruebas de rendimiento académico	176
Cuestionarios	177
Manuales	177
Agradecimientos	178
RECUADROS	
2.1 Currículo de matemáticas de Papúa Nueva Guinea	13
2.2 Currículo de lengua (inglés) de Nueva Zelanda	14
2.3 Ejemplos de ítems de opción múltiple	21
2.4 Ejemplo de ítem de respuesta cerrada	22
2.5 Ejemplos de ítems de respuesta corta abierta	22
2.6 Ejemplo de instrucciones para la escritura de un ensayo	23
3.1 Ejemplo de material de estímulo irrelevante	36
3.2 Ejemplo de un ítem con información inexacta o engañosa	37
3.3 Ejemplo de un ítem de opción múltiple	39
3.4 Puntuación en oraciones completas	40
3.5 Puntuación en una lista	40
3.6 Minimizar la cantidad de texto que debe leerse	41
3.7 Ítem con un enunciado negativo	41
3.8 Distractores mal emparejados	42
3.9 Uso de pares en los distractores	42
3.10 Ítem de respuesta abierta confuso con indicaciones poco claras	45
3.11 Un buen ejemplo de un ítem con respuesta cerrada	46
3.12 Ítem de crédito parcial	47
3.13 Ejemplo de un ítem de respuesta abierta con guía de corrección	48
3.14 Ejemplo de un ítem de respuesta cerrada con guía de corrección	49
3.15 Uso de ilustraciones para reducir la cantidad de palabras	54
3.16 Uso de imágenes simples	55

3.17 Gráficos con etiquetas claras	55
3.18 Mapas con etiquetas claras	56
3.19 Dejar espacios en el material de estímulo	57
3.20 Ejemplo de hoja de estilo para los redactores de ítems	61
4.1 Ejemplo de pregunta de respuesta múltiple y formato abierto	81
4.2 Ejemplo de hoja de captura de datos para una prueba previa	84
6.1 Ejemplo de una portada de pruebas	97
8.1 Especificaciones técnicas para el cuestionario sobre valores y actitudes	118
9.1 Correspondencia deficiente entre las casillas y las categorías de respuesta	130
9.2 Mejor correspondencia de casillas y categorías de respuesta	131
10.1 Ejemplo de codificación en escala de grises	134
10.2 Ejemplo de tratamiento de los ítems como categorías individuales para la captura de datos	134
12.1 Instrucciones del manual de administración	147
12.2 Información para docentes y directores	148
12.3 Administración de los ítems de práctica	149
13.1 Lista de verificación de administración: Un ejemplo de Filipinas	157

FIGURAS

1.1 Organigrama de una evaluación nacional	4
1.2 Resumen de actividades de una evaluación nacional	5
4.1 Ejemplo de vinculación circular de ítems	74
4.2 Modelo para ítems de vinculación vertical	75
C.1 Guía para acceder a los materiales en el CD sobre pruebas, cuestionarios y manuales	176

TABLAS

1.1 Etapas del desarrollo de la prueba y el diseño del cuestionario de evaluación nacional	6
2.1 Especificaciones técnicas para una prueba de matemáticas en el ciclo medio de la escuela primaria	15
2.2 Especificaciones técnicas para TIMSS matemáticas, 3.º y 4.º grado	16
2.3 Especificaciones técnicas de los contenidos de matemáticas para Papúa Nueva Guinea	17

2.4	Ventajas y limitaciones de los formatos de los ítems	26
2.5	Prueba de matemáticas de Papúa Nueva Guinea - Formatos de ítems	27
4.1	Ítems de enlace en dos unidades de lectura	76
4.2	Parte de una hoja de cálculo para controlar los ítems en los distintos formularios	77
5.1	Ejemplo de resultado del análisis de un ítem de respuesta múltiple	90
5.2	Ejemplo de resultado del análisis de un ítem de respuesta abierta y de crédito parcial	92
8.1	Partes del desarrollo del cuestionario	114
8.2	Funciones de la comprensión lectora en un estudio internacional: Ponderaciones utilizadas para crear dos nuevas variables, “leer con fines utilitarios” y “leer por placer”	120



PRÓLOGO

La calidad de cualquier ejercicio de evaluación educativa depende de la calidad de los instrumentos que se hayan empleado. De hecho, si el diseño de estos instrumentos es deficiente, la evaluación puede ser una pérdida de tiempo y dinero. *Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico* es el segundo de cinco libros dentro de la serie *Evaluaciones nacionales del rendimiento académico*. En él se describe de qué manera desarrollar instrumentos técnicamente confiables para una evaluación nacional del rendimiento en educación, y se concentra principalmente en la puesta en práctica de esta tarea en países en desarrollo. El volumen 1 de esta serie describe los principales objetivos y las características sobresalientes de las evaluaciones nacionales y está destinado principalmente a los responsables de las políticas y la toma de decisiones en el campo de la educación. Este segundo libro y la mayor parte de los posteriores dentro de esta serie brindan detalles paso a paso sobre el diseño, la implementación, el análisis y la generación de informes de las evaluaciones nacionales y están dirigidos principalmente a los equipos de evaluación nacional.

Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico aborda el diseño de dos tipos de instrumentos de recopilación de datos: las pruebas de rendimiento estudiantil y los cuestionarios de contexto. La primera parte trata del desarrollo de un marco de evaluación y de unas especificaciones técnicas para las

pruebas, la redacción de los ítems, las pruebas previas y la estructura final de las pruebas. La segunda parte determina las etapas y actividades correspondientes a la preparación de cuestionarios de contexto, que se utilizan para recopilar información sobre los estudiantes, los docentes, los directores de escuela o los padres acerca de las variables que podrían ayudar a explicar las diferencias en el rendimiento de los estudiantes en las pruebas de rendimiento académico. La tercera describe cómo diseñar un manual para la administración de las pruebas que sirva para asegurar que todos los estudiantes rindan en condiciones estandarizadas. El disco compacto (CD) que acompaña este libro ofrece buenos ejemplos de diseño de ítems de las pruebas, ítems de los cuestionarios y manuales de administración de las pruebas tomados de evaluaciones nacionales e internacionales y está destinado a brindar casos prácticos de la variedad de métodos que los equipos de evaluación han adoptado para el diseño de estos instrumentos.

El volumen 3 de esta serie se concentra en cuestiones prácticas que deben encararse al implementar un programa de evaluación nacional a gran escala, como son la logística, el muestreo y la limpieza y gestión de los datos. El volumen 4 trata cómo generar datos sobre los ítems y los puntajes de las pruebas y cómo relacionar esos puntajes con otros factores educativos. Por último, el volumen 5 se refiere a la forma de redactar los informes basados en las conclusiones de la evaluación nacional y cómo utilizar los resultados para mejorar la calidad de las políticas educativas.

A medida que los lectores avancen en este volumen, resultará evidente que el desarrollo de los instrumentos de evaluación es un ejercicio complejo, que demanda mucho tiempo y exige una cantidad considerable de conocimientos, habilidades y recursos. Al mismo tiempo, la experiencia ha demostrado que el beneficio que se deriva de contar con instrumentos bien diseñados puede ser sustancial en términos de calidad de la información obtenida sobre los niveles de rendimiento de los estudiantes y sobre los factores escolares y no escolares que podrían ayudar a mejorar esos niveles de rendimiento. La buena calidad de los instrumentos puede aumentar la confianza de los responsables políticos y otros actores en las conclusiones. También puede aumentar la probabilidad de que los responsables políticos utilicen los resultados de la evaluación nacional para el desarrollo de

planes y programas bien fundados destinados a mejorar la calidad de la educación. Si los resultados de las pruebas y los cuestionarios logran estos efectos, el tiempo y esfuerzo invertidos en su desarrollo estarán más que justificados.

Marguerite Clarke

Especialista Superior en Educación Banco Mundial



ACERCA DE LOS AUTORES Y EDITORES

AUTORES

Prue Anderson es investigadora superior en el Consejo Australiano de Investigación Educativa. Ha desarrollado material de evaluación de lectura para los programas de pruebas por nivel para alumnos de escuela primaria y primer ciclo de secundaria. Ha trabajado en programas de monitoreo educativo en Australia, Brunéi Darussalam, Papúa Nueva Guinea y Filipinas. Actualmente, se desempeña como gerente de proyecto del programa para la Evaluación Internacional de Escuelas (ISA). Otras áreas de interés profesional incluyen (a) correlación de datos de evaluación con marcos y resultados curriculares y (b) medición de resultados sociales de la escolarización.

George Morgan es asesor educativo. Ha sido investigador superior de la División de Medidas y estuvo a cargo del Grupo de Desarrollo de Pruebas en Ciencias y Matemáticas en el Consejo Australiano de Investigación Educativa durante casi 30 años. Ha desarrollado planes de estudio para ciencias y matemáticas y material de evaluación para todos los niveles educativos y trabajado en programas de pruebas a gran escala. Hace muy poco, participó de manera activa en los proyectos de evaluación en Camboya, Timor Oriental, República Democrática Popular Lao, Papúa Nueva Guinea y Samoa.

EDITORES

Vincent Greaney es asesor educativo. Ha sido jefe de especialistas en educación en el Banco Mundial y ha trabajado en varios países de África, Asia y Oriente Medio. Ex docente, investigador en el Centro de Investigación Educativa del St. Patrick's College, Dublín; profesor visitante becario Fulbright en la Universidad del Oeste de Michigan, Kalamazoo; miembro de la Galería de Honor de la Lectura de la Asociación Internacional de Lectura (IRA). Sus áreas de investigación más importantes son evaluación, formación docente, lectura y promoción de la cohesión social mediante la reforma de los libros de texto.

Thomas Kellaghan es director del Centro de Investigación Educativa del St. Patrick's College, Dublín, y miembro numerario de la Academia Internacional de Educación. Ha trabajado en la Universidad de Ibadán, en Nigeria, y en la Queen's University de Belfast. Su investigación se centra en las evaluaciones y los exámenes, la desventaja educativa y las relaciones entre el hogar y la escuela. Se desempeñó como presidente de la Asociación Internacional de Evaluación Educativa entre 1997 y 2001. Ha trabajado en temas relacionados con las evaluaciones en África, Asia, América Latina y Oriente Medio.



AGRADECIMIENTOS

Un equipo liderado por Vincent Greaney (asesor, Red sobre Desarrollo Humano, Grupo de Educación, Banco Mundial) y Thomas Kellaghan (Centro de Investigación Educativa, St. Patrick's College, Dublín) ha preparado la serie de libros *Evaluaciones nacionales del rendimiento académico*, de la cual este es el segundo volumen. Han colaborado también en esta serie: Sylvia Acana (Junta Nacional de Exámenes de Uganda), Prue Anderson (Consejo Australiano de Investigación Educativa), Fernando Cartwright (Consejo Canadiense sobre Aprendizaje), Jean Dumais (Dirección General de Estadísticas de Canadá), Chris Freeman (Consejo Australiano de Investigación Educativa), Hew Gough (Dirección General de Estadísticas de Canadá), Sara Howie (Universidad de Pretoria), George Morgan (Consejo Australiano de Investigación Educativa), T. Scott Murray (Instituto de Estadística de la UNESCO) y Gerry Shiel (Centro de Investigación Educativa, St. Patrick's College, Dublín). El trabajo se llevó a cabo bajo la dirección general de Ruth Kagia, directora de educación, y Robin Horn, gerente de la Red sobre Desarrollo Humano del Grupo de Educación, ambos del Banco Mundial. Robert Prouty inició el proyecto y lo dirigió hasta agosto de 2007. Marguerite Clarke se hizo cargo del proyecto en las etapas posteriores de revisión y publicación.

Agradecemos la colaboración del grupo de revisión: Al Beaton (Boston College), Irwin Kirsch (Education Testing Service [Servicio de

Pruebas Educativas], ETS), y Benoit Millot (Banco Mundial). A Helen Abadzi, Regina Bendokat, Marguerite Clarke, Robin Horn, Elizabeth King, Maureen Lewis, Harry Patrinos, Carlos Rojas, Jee-Peng Tan, Eduardo Vélez y Raisa Venalainen, por su aporte de comentarios adicionales de gran utilidad.

Recibimos el valioso aporte y respaldo de Carly Cheevers, David Harding, Aidan Mulkeen, Aleksandra Sawicka, Thi Tran, Hilary Walshe y Hans Wagemaker.

Queremos agradecer a las siguientes instituciones el permiso otorgado para reproducir su material en el cuerpo de este volumen y en el disco compacto que acompaña esta obra: Consejo Australiano de Investigación Educativa; Centro de Investigación Educativa, Dublín; Asociación Internacional para la Evaluación del Rendimiento Educativo; Departamento de Educación de Massachusetts; Centro Nacional de Estadística Educativa de los Estados Unidos. Departamento de Educación; Organización para la Cooperación y el Desarrollo Económicos; y Departamento de Educación de Papúa Nueva Guinea.

Mary Fisk y Paola Scalabrin, de la Oficina del Editor del Banco Mundial, coordinaron el diseño, la edición y la producción del libro.

El Fondo Fiduciario de Irlanda para la Educación, el Programa de Asociación Banco-Países Bajos (BNPP), el Centro de Investigación Educativa, Dublín, y el Consejo Australiano de Investigación Educativa brindaron su generoso apoyo para la preparación y publicación de esta serie.



SIGLAS

CD	disco compacto
CDN	comité director nacional
ID	número de identificación
MdE	ministerio de educación
NAEP	Evaluación Nacional del Progreso Educativo (Estados Unidos)
PIRLS	Estudio sobre el Progreso Internacional de la Competencia en Lectura
PISA	Programa para la Evaluación Internacional de Alumnos
SSD	supervisor superior de distrito
STBA	asignación de cuadernillos de la prueba de los estudiantes
TCP	teoría clásica de las pruebas
TIMSS	Tercer Estudio Internacional de Matemáticas y Ciencias o Estudio Internacional de Tendencias en Matemáticas y Ciencias
TRI	teoría de respuesta al ítem

PARTE

1

ELABORACIÓN DE LAS PRUEBAS

INTRODUCCIÓN

Las actividades de una evaluación nacional son muchas, desde el momento en que alguien decide llevarla a cabo hasta el momento en que se lee el informe con sus resultados. Cada uno de los libros en esta serie de cinco volúmenes, intitulada *Evaluaciones nacionales del rendimiento académico*, describe algunas de las actividades de una evaluación nacional, enfatizando la puesta en práctica de dicha evaluación en los países en desarrollo. Es probable que gran parte de la tecnología necesaria para llevar a cabo una evaluación nacional satisfactoria no exista en los países con poca tradición en investigación empírica en educación. Por consiguiente, en esta serie se ha hecho el esfuerzo de describir al detalle las actividades de una evaluación y, cuando corresponde, de ayudar a los lectores (que podemos suponer tienen responsabilidad en al menos algunos de los aspectos de una evaluación) a comprender por qué son necesarias esas actividades.

El ministerio de educación (MdE), o el comité director nacional (CDN) designado, en general tendrá la responsabilidad global de guiar y respaldar la evaluación nacional. La mayor parte del trabajo será llevado a cabo por una agencia implementadora, bajo la supervisión del MdE o del CDN. A su vez, dicha agencia supervisará la tarea del encargado del desarrollo de la prueba, de los especialistas en las diversas materias y de los analistas de estadísticas, y se encargará de la

logística para la ejecución de la evaluación nacional. Este libro, *Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico*, trata principalmente de las actividades del encargado del desarrollo de la prueba y de los especialistas en las diversas materias, así como también el desarrollo de pruebas previas (véase la figura 1.1). Otro de los temas que se muestran en la figura 1.1, tales como el muestreo; los aspectos logísticos de la evaluación, entre otros la comunicación con las escuelas; y la captura y depuración de los datos, se tratan en el volumen 3 de esta serie: *Implementación de una evaluación nacional del rendimiento académico*. El volumen 4, *Análisis de los datos de una evaluación nacional del rendimiento académico*, está dedicado al análisis estadístico.

El diagrama de flujo que aparece en la figura 1.2 resume los distintos pasos de una evaluación nacional. Muchos de esos pasos están descritos en este libro; los recuadros o las actividades sombreados se

FIGURA 1.1

Organigrama de una evaluación nacional

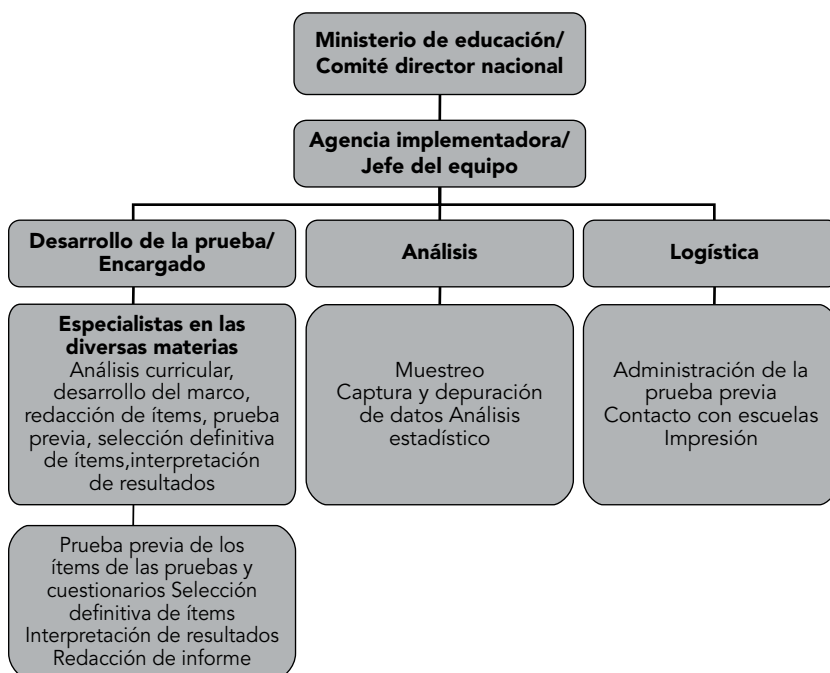


TABLA 1.1
Etapas del desarrollo de la prueba y el diseño del cuestionario de evaluación nacional

Etapa	Actividad	Tiempo aproximado	Personas que participan
1	Preparar el marco de evaluación; clarificar el propósito de las evaluaciones, las pruebas y los cuestionarios nacionales; y seleccionar la población objetivo.	4 semanas	MdE o CDN y agencia implementadora, en especial el líder de grupo, el encargado del desarrollo de la prueba, las partes interesadas clave y los responsables de las políticas.
2	Diseñar las especificaciones técnicas, o tabla de especificaciones, y consultar a todas las partes para su aprobación.	4 a 6 semanas	MdE o CDN, agencia implementadora, encargado del desarrollo de la prueba, grupos de especialistas, docentes experimentados, especialistas en materias, analistas de datos, redactores de ítems experimentados, partes interesadas clave y responsables de las políticas.
	Redactar los ítems de la prueba y del cuestionario. ^a	12 a 14 semanas (entre 20 y 30 ítems por redactor por semana)	Encargado del desarrollo de la prueba, especialistas en las materias, redactores de ítems, y partes interesadas clave.
	Redactar las pruebas previas y los borradores de los cuestionarios.	4 semanas	Jefe del equipo, encargado del desarrollo de la prueba, redactores de ítems, profesionales de diseño y diagramación y correctores.
	Imprimir las pruebas previas y los borradores de cuestionarios.	2 semanas	Agencia implementadora, jefe del equipo, encargado del desarrollo de la prueba y redactores de ítems.
	Preparar los paquetes de pruebas previas y borradores de cuestionarios y distribuirlos.	2 a 3 semanas	Agencia implementadora.

3	Administrar las pruebas previas y los borradores de cuestionarios en las escuelas.	2 a 3 semanas	Agencia implementadora y examinadores.
	Calificar los ítems manualmente (si fuera necesario).	2 semanas	Jefe del equipo, encargado del desarrollo de la prueba, especialistas en las diversas materias y redactores de ítems.
	Ingresar los datos de las pruebas previas.	1 semana	Analista de datos de la agencia implementadora y personal de captura de datos.
4	Analizar los datos de las pruebas previas y los cuestionarios.	2 semanas	Agencia implementadora, analistas de datos, redactores de ítems y encargado del desarrollo de la prueba.
	Seleccionar ítems para las pruebas y cuestionarios.	2 semanas	Encargado del desarrollo de la prueba, analistas, redactores de ítems y actores clave.
5	Producir las pruebas, los cuestionarios y los manuales de administración definitivos.	2 semanas	Agencia implementadora, encargado del desarrollo de la prueba, profesionales de diseño y diagramación, correctores y redactores de ítems.
	Imprimir las pruebas y cuestionarios.	4 semanas	Agencia implementadora, jefe del equipo y encargado del desarrollo de la prueba.
	Preparar los paquetes de pruebas y cuestionarios y distribuirlos.	2 a 3 semanas (según la distancia y accesibilidad)	Agencia implementadora y encargado del desarrollo de la prueba.

(continúa)

TABLA 1.1 (continúa)

Etapa	Actividad	Tiempo aproximado	Personas que participan
6	Administrar las pruebas y los cuestionarios en las escuelas. Calificar los ítems manualmente (si fuera necesario).	3 a 4 semanas 3 a 4 semanas	Agencia implementadora, encargado del desarrollo de la prueba y examinadores. Encargado del desarrollo de la prueba y redactores de ítems.
7	Ingresar y depurar datos. Analizar los datos.	4 a 6 semanas 2 a 3 semanas	Analista de datos y personal de captura de datos. Analista de datos, redactores de ítems y encargado del desarrollo de la prueba.
8	Redactar los informes definitivos.	4 a 5 semanas	Analista de datos, redactores de ítems y encargado del desarrollo de la prueba.

a. Se necesitará tiempo adicional si se deben traducir los ítems a otros idiomas.

referen a los aspectos de la evaluación que reciben una mayor atención en esta obra. El libro también incluye una cantidad de notas o comentarios comunes para más de un aspecto de la evaluación; están reiterados para facilitar la lectura de aquellos interesados en un único aspecto de la evaluación nacional.

En la tabla 1.1 se brinda información adicional sobre el desarrollo de la prueba y el esquema del cuestionario. En esta tabla se describe un proceso de elaboración de pruebas y cuestionarios de logros académicos que comprende ocho etapas e indica quiénes son las personas responsables de los componentes de cada etapa.

El disco compacto (CD) que acompaña este libro ofrece muchos ejemplos de ítems de las pruebas, ítems de los cuestionarios y manuales de administración de las pruebas. El Apéndice C presenta más detalles del contenido del CD. El presente material ha sido extraído de evaluaciones nacionales e internacionales, y pretende familiarizar a los equipos de evaluación nacionales con los ítems y tipos de ítem de varias áreas curriculares y con los cuestionarios diseñados para alumnos, maestros, escuelas, directores y padres.

DESARROLLO DE UN MARCO DE EVALUACIÓN

Contar con un marco que brinde un esquema o plan general para guiar el desarrollo de las pruebas de evaluación, los cuestionarios y los procedimientos es crucial para determinar los contenidos de esa evaluación. (Linn y Dunbar 1992; Mullis et al. 2006). Dicho marco ayuda a generar una clara comprensión del constructo que se pretende evaluar (por ejemplo, el rendimiento en lectura o matemáticas) y los distintos procesos del mismo. Debe incluir una definición de lo que se está evaluando, determinar las características de las tareas que se emplearán en el desarrollo de la prueba y establecer las bases para interpretar los resultados (Kirsch 2001; Messick 1987). Un marco puede ayudar a explicar el propósito de la evaluación. Puede facilitar los debates y la toma de decisiones entre los distintos actores educativos aclarando los conceptos clave antes de que se inicie la evaluación. El marco también puede determinar las variables clave que probablemente tengan correlación con el puntaje de las pruebas y puede ayudar a asegurar que esas variables se incluyan en el diseño de la evaluación nacional.

Desde el principio, el comité director debe definir lo que se ha de medir. Con frecuencia, el documento del currículo nacional contendrá definiciones sobre las áreas temáticas clave. Las definiciones de lectura, por ejemplo, han variado en el tiempo y entre los distintos sistemas educativos. En algunos casos, ha sido equivalente a la capacidad de

pronunciar palabras. En otros, se refiere a la capacidad de reconocer palabras individuales y asignarles su significado. La lectura también ha sido definida como la capacidad de comprender o captar el significado de un texto. Las definiciones más recientes van más allá de las simples habilidades de decodificación e incluyen la capacidad de aplicar la información contenida en los textos así como también de comprenderlos. También reconocen que los estudiantes y los adultos leen con distintos propósitos, por ejemplo por placer o por información. Estas nuevas definiciones se reflejan en las pruebas al incluir distintas formas de textos, como por ejemplo, cuentos, fragmentos de artículos de diarios, anuncios comerciales, señales y gráficas.

En el desarrollo de las pruebas debe quedar claro el propósito para el cual se recopilan los datos. Es de gran importancia que en las etapas iniciales se realicen consultas con actores y grupos de expertos clave a fin de determinar claramente el propósito de la evaluación nacional y, por ende, qué debe evaluar la prueba, quiénes deben ser evaluados, cuándo deben evaluarse y en qué idioma deben impartirse las pruebas. Los expertos en currículo deben ser parte de estas decisiones, junto con los responsables de las políticas y de la gestión educativa, ya que son los que estarán en condiciones de emplear los resultados de la evaluación como base para las políticas educativas, para la asignación de recursos y para la implementación de reformas.

Las evaluaciones nacionales pueden ser herramientas poderosas para medir la eficiencia de algunos aspectos del currículo. Un buen diseño de la evaluación también puede reforzar las metas del currículo simulando los tipos de habilidades y saberes que los alumnos deben demostrar. Estos tipos de habilidades y los contextos en los cuales se evalúan deberían respaldar las metas de las políticas educativas globales en las áreas clave del aprendizaje. Los ejemplos de algunos de los contextos globales de las evaluaciones nacionales en los recuadros 2.1 y 2.2 reflejan varias prioridades educativas.

DOCUMENTO DE ESPECIFICACIONES TÉCNICAS DE LA PRUEBA O TABLA DE ESPECIFICACIONES

El documento de especificaciones técnicas de la prueba, o tabla de especificaciones, es el documento esencial que guía su desarrollo, su

RECUADRO 2.1**Currículo de matemáticas de Papúa Nueva Guinea**

El currículo 2003 para las escuelas primarias de Papúa Nueva Guinea a nivel elemental de matemáticas en el contexto cultural actual se propone con la siguiente justificación global:

Todos los ciudadanos tienen derecho a participar en el desarrollo futuro de Papúa Nueva Guinea. Por esta razón, es necesario que los alumnos desarrollen con solidez conocimientos, habilidades y comprensión en matemáticas... los alumnos de [nivel] elemental serán capaces de relacionar los nuevos conceptos matemáticos de los cinco ejes temáticos dentro de este programa con su conocimiento actual de su contexto cultural, de modo tal que puedan aplicar con seguridad las matemáticas en su vida diaria. El curso de matemáticas en el contexto cultural de primaria brinda muchas oportunidades para el aprendizaje pertinente y orientado hacia los principios de la vida en el hogar. (Papúa Nueva Guinea, Departamento de Educación 2003: 2)

Esta justificación (junto con los documentos de la política ministerial y con una importante reestructuración de la educación primaria y preprimaria) está dirigida a integrar las matemáticas de la primaria en la cultura de la comunidad de los alumnos. Las reformas han establecido la prioridad en la integración de las matemáticas del nivel primario con la cultura local y en la aplicación de los conocimientos matemáticos en la vida cotidiana. Una evaluación nacional reciente desarrollada para el seguimiento del rendimiento de los alumnos hizo hincapié en el empleo de contextos realistas para los cuestionarios y para la evaluación de las habilidades y conocimientos que tengan aplicaciones prácticas.

análisis y sus reportes. Describe los datos que deben ser recopilados, define la extensión de la prueba y especifica la proporción de ítems que abordarán los distintos aspectos del currículo dentro de una prueba. Unas buenas especificaciones técnicas deberán señalar lo siguiente:

- La proporción de ítems dentro de la prueba, en su forma final, que abordan cada una de las áreas del currículo (por ejemplo, matemáticas, lengua, ciencias).
- La proporción de ítems dentro de un área del currículo que evalúan diferentes habilidades (por ejemplo, en matemáticas: números,

RECUADRO 2.2**Currículo de lengua (inglés) de Nueva Zelanda**

Los objetivos generales del currículo de lengua (inglés) de Nueva Zelanda establecen:

Los alumnos deben ser capaces de interesarse en el lenguaje y disfrutar de él en todas sus variedades así como comprender, responder y utilizar de manera eficaz el lenguaje oral, escrito y visual en una diversidad de contextos. (Nueva Zelanda, Ministerio de Educación 2002: 9)

Estos objetivos destacan la importancia del interés y el placer en la lectura y en la comprensión de una amplia variedad de textos. El recurso a textos y a tareas agradables y significativas es un aspecto clave que debe tomarse en consideración en las evaluaciones globales en la materia inglés. El énfasis en el idioma en todas sus formas refleja un fuerte compromiso con el reconocimiento y la valoración de la cultura de la lengua oral de los alumnos maoríes, así como también de las formas escritas del inglés. Varias evaluaciones nacionales reflejan estos objetivos.

medidas, espacio y patrones; en escritura: ideas, conocimiento del contenido, estructura, estilo, vocabulario, ortografía y gramática).

- La proporción de ítems que abordan distintas habilidades del proceso cognitivo (como por ejemplo conocimiento o memoria, interpretación o reflexiones).
- La proporción de ítems de opción múltiple y de respuesta abierta.
- La proporción de ítems dedicados a textos de estímulo en distintos tipos de lectura (como por ejemplo textos narrativos, expositivos, instructivos y argumentativos) o en matemáticas (tales como tablas, gráficas y diagramas).

Las especificaciones técnicas de la prueba que se muestran en la tabla 2.1 se basan en un currículo de matemática para los grados intermedios de la escuela primaria. Se diseñaron subpruebas independientes para medir las habilidades de los alumnos para realizar cálculos básicos, comprender conceptos matemáticos y resolver problemas. Por ejemplo, la celda formada por la intersección del área de contenido “fracciones” y la de comportamiento intelectual “capacidad de resolución de problemas rutinarios” representa el objetivo “capacidad de resolver problemas rutinarios con fracciones”.

TABLA 2.1
Especificaciones técnicas para una prueba de matemáticas en el ciclo medio de la escuela primaria

Áreas de contenido	Comportamiento intelectual												
	Cálculo			Conceptos					Resolución de problemas				
	Conocimiento de términos y hechos (A1)	Capacidad de realizar operaciones (A1)	Total	Comprensión de conceptos matemáticos (B1)	Comprensión de principios matemáticos (B2)	Comprensión de estructura matemática (B3)	Capacidad de trasladar elementos de una forma a otra (B4)	Capacidad de leer e interpretar gráficas y diagramas (B5)	Total	Capacidad de resolver problemas rutinarios (C1)	Capacidad de analizar y realizar comparaciones (C2)	Capacidad de resolver problemas no rutinarios (C3)	Total
1. Números enteros		25	26	1	4	7	2	4	18	14	2	2	18
2. Fracciones		4	4	4	1	2			7	5			5
3. Decimales		8	8			5		1	6	5			5
4. Medidas	2		2				3	2	5	3			3
5. Geometría			0	2	2				4				0
6. Cuadros y gráficas			0						0		4		4
Total General	3	37	40	7	7	14	5	7	40	27	6	2	35
Total General													115

Fuente: Centro para la Investigación Educativa 1978: 44.

Un comité de especialistas en la materia, que incluyó a los docentes, decidió dedicar cinco ítems a ese objetivo. La celda que contiene los ítems que evalúan la capacidad de realizar operaciones con números enteros recibió la ponderación más alta (25 ítems). Muchas celdas no tuvieron ningún ítem. La importancia relativa atribuida a cada objetivo se usó como guía para el desarrollo de la prueba y la posterior redacción de la versión final de la prueba.

Las especificaciones técnicas para matemáticas tomadas a partir del Estudio Internacional de Tendencias en Matemáticas y Ciencias (prueba TIMSS)¹ en la tabla 2.2 definen los formatos de los ítems y los procesos cognitivos que se abordarán de manera algo diferente en el tercer y cuarto grado.

Sin duda, las especificaciones técnicas varían de acuerdo con la comprensión del constructo que se pretende medir y con el propósito de la evaluación. Quienes participan en el desarrollo de la prueba deben comprender y aprobar las implicaciones de unas especificaciones técnicas en términos de qué debe evaluarse y qué debe dejarse fuera de la prueba.

Debido a las limitaciones de tiempo y de recursos, no es posible evaluar cada subje de un área del currículo o todos los temas incluidos en un programa de estudios. Los ítems de la prueba siempre deben estar dirigidos a las habilidades centrales. Deberá consultarse a los expertos en currículo o en las materias para determinar cuáles son las habilidades centrales. La importancia que las especificaciones

TABLA 2.2
Especificaciones técnicas para TIMSS matemáticas, 3.º y 4.º grado

Expectativas de rendimiento	Número total de ítems ^a	Ítems de opción múltiple	Ítems de respuestas cortas	Ítems de respuesta amplia
Conocimiento	42	35	7	0
Ejecución de procedimientos rutinarios	16	13	3	0
Empleo de procedimientos complejos	24	21	2	1
Resolución de problemas	20	10	3	7

Fuente: IEA, <http://timss.bc.edu/timss1995i/TIMSSPDF/AMitems.pdf>.

a. La cantidad de ítems refleja la combinación total de ítems que se utilizó para formar 26 grupos de pruebas en ocho cuadernillos de prueba diferentes. No se requirió que ninguno de los alumnos rindiera la prueba completa.

técnicas de la evaluación nacional asignen a los subejos o dominios de un currículo también depende de la manera en que se vaya a comunicar los datos de la prueba (por puntaje general o por dominio del currículo). Deberá consultarse a los responsables de las políticas educativas para establecer cómo prefieren que se comuniquen los datos de las pruebas.

Si los datos se han de transmitir como un único puntaje general para cada estudiante en un área del currículo, por ejemplo en matemáticas, entonces se requieren al menos 25 o 30 ítems. Puede que se requieran datos en un número de ítems un poco más pequeño para informar acerca de un subejo, como por ejemplo la comprensión de los alumnos del espacio o de la resolución de problemas en una prueba de matemáticas.

La tabla 2.3 brinda un ejemplo del cuadro de especificaciones de contenido para una prueba de matemáticas para alumnos de 3.º, 5.º y 8.º grado en Papúa Nueva Guinea. Observe que en la prueba para 3.º grado, más del 80 por ciento de los ítems están dedicados a números y su aplicación, espacio y forma y medidas. Solo cuatro ítems se refieren a patrones y probabilidad. Para el 8.º grado los ítems se distribuyen de manera más pareja entre los distintos subejos.

Las especificaciones técnicas en general se basan en un currículo sugerido (o pretendido). Sin embargo, si no se consideran tanto el currículo implementado (lo que los docentes enseñan) como el currículo alcanzado (lo que los alumnos han aprendido) puede que la prueba resulte demasiado difícil o demasiado fácil. Por lo tanto, no proporcionará una descripción significativa del grado de rendimiento de los alumnos de la población objetivo. Si la mayor parte de los alumnos no logra responder correctamente los ítems, la prueba no

TABLA 2.3
Especificaciones técnicas de los contenidos de matemáticas para Papúa Nueva Guinea

Grado	Números y aplicación	Forma y espacio	Medidas	Probabilidad	Patrones y álgebra ^a	Total de ítems
3.º grado	10	7	4	2	2	25
5.º grado	10	10	7	4	4	35
8.º grado	10	10	8	6	6	40

Fuente: material inédito, Departamento de Educación de Papúa Nueva Guinea.

a. Álgebra aplica solamente al octavo grado.

muestra si estos alumnos están capacitados para demostrar habilidades con un nivel de dificultad apenas inferior o muy por debajo del nivel de los ítems comprendidos en la prueba. Del mismo modo, si la mayoría de los alumnos responden todos los ítems correctamente, la prueba no indica si son capaces de demostrar habilidades con una dificultad apenas superior o muy por encima de la de los ítems de la prueba. Los datos provenientes de las pruebas demasiado difíciles o demasiado fáciles son de escasa utilidad para los responsables políticos, para las escuelas o para los docentes.

La dificultad de la prueba depende del fin que persigue. Si el propósito es monitorear el desempeño de todos los alumnos en la población objetivo, entonces la distribución de la dificultad de los ítems de la prueba deberá concordar con la distribución de logros de dicha población. Como regla general, dos tercios de la prueba deben corresponder a ítems en los cuales dos tercios de la población tengan entre un 30 y un 70 por ciento de probabilidades de responder correctamente. (En promedio, la probabilidad debe ser del 50 por ciento para ayudar a maximizar la variación en los puntajes de los alumnos en estas pruebas). El restante tercio de la prueba deberá dividirse de forma equitativa entre los ítems en los que es probable que más del 70 por ciento de los alumnos que realizan la prueba respondan correctamente e ítems donde es probable que menos del 30 por ciento lo haga. Si bien la sensibilidad a los rendimientos de los alumnos que reflejan estas cifras es importante, no debe generar la exclusión de áreas importantes del currículo simplemente porque el desempeño de los alumnos en estas áreas sea muy deficiente o muy bueno. La idoneidad de los ítems deberá establecerse en un programa de pruebas previas, en las cuales esos ítems se emplean con alumnos con características similares a los de la población objetivo de la evaluación nacional.

Algunas evaluaciones nacionales establecen niveles de rendimiento o desempeño que se basan en una norma predefinida e identifican a los alumnos que han alcanzado ese nivel. Si ese nivel es muy alto, la prueba identificará a un pequeño número de alumnos que demuestran dicho nivel de habilidades, pero presentará muy poca información acerca del nivel de rendimiento del resto de la población, excepto que el nivel está por debajo de la norma. Si esa norma es muy baja, la prueba identificará a la gran masa de alumnos que demuestran este nivel de destreza, pero

presentará muy poca información acerca de niveles superiores de destrezas que también podrían haber adquirido estos alumnos.

VALIDEZ

La validez es un constructo amplio que implica realizar interpretaciones y usos adecuados de los resultados o de la información obtenida de la prueba (Messick 1989). Una de las facetas de la validez es el grado de representatividad del contenido de la prueba en lo que se refiere al currículo o al constructo que se pretende medir. El encargado del desarrollo de la prueba es responsable de coordinar un grupo de referencia de especialistas en las diversas materias, por ejemplo especialistas en currículo, a fin de garantizar que los ítems representen un muestreo adecuado del currículo o del constructo. El grupo de expertos no debe incluir a los redactores de los ítems. En esta instancia, la validez es una cuestión de criterio y no de estadística. El grupo de expertos deberá determinar si la prueba representa una cobertura adecuada de la materia específica (por ejemplo matemáticas de 4.º grado) y deberá considerar si el desempeño en la prueba brinda evidencias adecuadas del rendimiento del estudiante en esa área temática.

IDIOMA DE LA PRUEBA

El marco de la prueba deberá clarificar y justificar el o los idiomas que se utilizarán en las pruebas de evaluación nacionales. El idioma de una prueba generalmente es el medio de instrucción. La traducción de los ítems de las pruebas, en aquellos casos en los que la educación se imparte en distintas lenguas, en general resulta costosa y demanda mucho tiempo. Las versiones traducidas de las pruebas deben ser lo más equivalentes posible si se pretende usar los datos con fines comparativos. A continuación se presentan algunas cuestiones que se deben considerar al incluir en la prueba un área particular del currículo utilizando más de un idioma.

- Si los recursos son limitados, puede que sea preferible evaluar a los alumnos de mayor edad en un idioma común de instrucción.

- Si se reducen las palabras utilizadas en los ítems de la prueba hasta el mínimo indispensable se pueden reducir los costos de traducción, pero generalmente también se descontextualiza el ítem, y por consiguiente se vuelve menos auténtico.
- Puede que sea preferible excluir a algunos alumnos de la población objetivo de una evaluación nacional en lugar de intentar adaptarla a todos los grupos lingüísticos.
- Algunas veces el idioma de instrucción pretendido no es la lengua en la que en realidad se enseña. En tales casos, los ítems de la prueba de evaluación nacional podrían utilizar la lengua real de enseñanza.
- En el caso de los alumnos de menor edad, especialmente si el idioma de enseñanza no es su primera lengua, la prueba se puede realizar de forma oral. El examinador lee cada una de las preguntas en voz alta o las graba y les da a los alumnos un tiempo suficiente para que respondan. Este método de ayuda puede resultar particularmente adecuado para las pruebas de matemáticas y ciencias para alumnos de menor edad, que podrían demostrar un mayor grado de conocimientos de la materia en pruebas orales que en una prueba que exige que lean los ítems por sí solos. Las pruebas diseñadas para evaluar las habilidades de lectura independiente de cada uno de los alumnos, por supuesto, no deberán administrarse de forma oral.

FORMATO DEL ÍTEM

En las evaluaciones escritas en papel, los alumnos responden una serie de preguntas o entradas. Las respuestas escritas o dibujadas se usan como prueba de su nivel de conocimiento, competencia o comprensión. Existen cuatro formatos básicos para los ítems o maneras en que el alumno puede presentar sus respuestas:

- Opción múltiple
- Respuesta cerrada
- Respuesta corta abierta
- Ensayo o respuesta amplia

En los *ítems de opción múltiple* (véase el recuadro 2.3) el alumno debe seleccionar una de varias opciones (en general, 4). Las opciones pueden estar escritas en letras o como imágenes con leyendas. Pueden estar en una lista —una debajo de la otra—, en una fila horizontal o en dos columnas. Los alumnos indican sus respuestas rellenando un círculo, encerrando una letra o número en un círculo, o marcando un casillero para seleccionar una porción de texto o un diagrama. Los ítems de opción múltiple tienen una opción “correcta” inequívoca y varias opciones posibles pero incorrectas. El disco compacto (CD) que acompaña este libro contiene muchos ejemplos de ítems de opción múltiple para lengua, matemáticas y ciencias para alumnos de nivel primario y posprimario.

Las *preguntas de respuesta cerrada* (véase el recuadro 2.4) tienen una respuesta correcta que escribe el alumno. Generalmente las variaciones mínimas en la forma de expresar la respuesta son aceptables. Puede ser que los alumnos deban escribir una o dos palabras, subrayar una palabra o número dentro de un texto o tabla, dibujar una línea en una cuadrícula, o indicar un área en un diagrama. Es posible que en

RECUADRO 2.3

Ejemplos de ítems de opción múltiple

1. ¿Cuál de los siguientes elementos es más probable que se mida en mililitros?

- A. La cantidad de líquido que contiene una cuchara de té
- B. El peso (masa) de un alfiler
- C. La cantidad de gasolina que contiene un tanque
- D. El espesor de 10 hojas de papel

Fuente: IEA 1998, ítem de muestra.

Nota: A es la respuesta correcta.

2. Una botella de jugo de manzana cuesta \$1,95. El pan cuesta \$2,75. ¿Cuál de las siguientes opciones es lo mínimo que se necesita para comprar el jugo de manzana y el pan?

- \$3,75
 \$4,00
 \$4,80
 \$5,00

Fuente: Consejo Australiano de Investigación Educativa, sin fecha, ítem de muestra.

Nota: \$4,80 es la respuesta correcta.

RECUADRO 2.4**Ejemplo de ítem de respuesta cerrada**

En el siguiente enunciado numérico:

$$2000 + \underline{\quad} + 30 + 9 = 2739$$

¿Qué número va en $\underline{\quad}$ para que el enunciado sea correcto?

Respuesta: $\underline{\hspace{2cm}}$

Fuente: IEA 1998, ítem de muestra.

RECUADRO 2.5**Ejemplos de ítems de respuesta corta abierta**

a. ¿Cómo se puede averiguar la edad de un árbol después de cortarlo?

b. Escribir un ejemplo de cómo las máquinas ayudan a las personas a hacer su trabajo.

Fuente: IEA 1998, ítem de muestra.

los ítems de respuesta cerrada, los alumnos también tengan que seleccionar varias opciones que cumplan con determinados criterios o hacer corresponder una serie de pares de oraciones o diagramas. (Véase en el CD, por ejemplo, los ítems 6, 9, 11 y 19 de la *Prueba de matemáticas 1990-2000 para 4.º grado de ENPE* y los ítems S011032 y S031053 de la *Prueba de ciencias para 4.º grado de la Encuesta TIMSS*).

En los ítems de *respuesta corta abierta* (véase el recuadro 2.5) los alumnos deben generar una respuesta para la cual pueden existir varias opciones diferentes pero correctas. En general, la respuesta se debe dar con una breve explicación, la demostración de un proceso o un dibujo detallado (más de uno o dos renglones). Es posible que el

alumno deba escribir una o dos oraciones, completar una serie de pasos o ecuaciones o completar varios aspectos de un dibujo, gráfica o diagrama. (Véase en el CD, por ejemplo, los ítems 6, 7 y 11 del “Apéndice B: Rescate del delfín”, *ítems de muestra de Lectura internacional para 4.º grado de la encuesta PIRLS 2006*).

En los ítems de *Ensayo* o *respuesta amplia* (véase el recuadro 2.6), los alumnos deben responder desarrollando un tema con respuestas

RECUADRO 2.6

Ejemplo de instrucciones para la escritura de un ensayo

Las instrucciones para la redacción de un ensayo pueden ser escritas o ilustradas. Es importante que los alumnos entiendan qué tipo de ensayo se les pide que hagan. Por ejemplo, puede ser que tengan que explicar sus ideas, expresar una opinión, escribir un texto persuasivo o narrar una historia. Esta información suele formar parte de las instrucciones que se dan durante la administración de la prueba.

La ilustración que se muestra en el siguiente ejemplo se usó en Papúa Nueva Guinea para evaluar la destreza de los alumnos en la redacción de un texto narrativo. El examinador les indicó que podían usar las ideas de la ilustración o las propias para escribir una historia sobre la caza.



Escribir una historia sobre la caza.

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

extensas, a veces de cierta complejidad. El texto de la respuesta puede ocupar una o más páginas o incluir diagramas. Existen muchas formas “correctas” de responder un ensayo o un ítem de respuesta amplia. (Véase, por ejemplo, la pregunta 9 del CD, “Un juez justo”, *ítems de lectura, Lectura Internacional del programa PISA 2000*, y los ítems 33, 35 y 39 de *Lectura principal ENPE, 1990-2006: 4.º grado*).

Los tres primeros formatos de ítems comúnmente son los más usados en evaluaciones nacionales, en parte por el costo que implica calificar manualmente una gran cantidad de ensayos de manera confiable. El documento marco de la prueba debe indicar los porcentajes estimados de los diferentes tipos de ítems de la prueba final. También debe incluir muestras de tipos de ítems para ayudar a los integrantes del comité director y a otros actores a familiarizarse con el enfoque de la evaluación.

La elección del formato del ítem y de su forma de puntuación influye significativamente en el costo total de la prueba. La calificación manual que exigen algunos ítems es más costosa, lleva más tiempo y por ende demora la publicación del informe. Se deben desarrollar guías para ese tipo de calificación y se deben contratar y capacitar correctores. Cuanto más complejas sean las guías de puntuación, mayores serán los costos. Los ítems de ensayos y de respuestas amplias suelen ser más costosos. La calificación de los ítems de opción múltiple suele ser menos costosa pero su elaboración es más cara comparada con otros tipos de ítems. A continuación se describen algunas cuestiones de costos a tener en cuenta al seleccionar el formato de ítem para la prueba.

En general, el software de análisis de datos califica los ítems de opción múltiple como correctos o incorrectos y no es necesario calificarlos antes del análisis. Solo hay que ingresar las respuestas de los alumnos en la computadora. Esto se puede hacer electrónicamente, mediante escaneo de las respuestas, o manualmente. El escaneo es el método menos costoso para pruebas a gran escala. Se necesitan equipos especiales y en algunos casos asistencia técnica de respaldo. Los ítems escaneados pueden limitarse a un formato particular de respuesta (por ejemplo, sombreado del círculo). Pero si el ingreso de datos es manual, se pueden incluir más estilos de opción múltiple (por ejemplo, encerrar palabras en un círculo, marcar casilleros o unir con líneas las opciones elegidas).

La calificación de los ítems de opción múltiple no debe quedar limitada a correctos o incorrectos sin un análisis previo de los datos, ya que se puede obtener información de diagnóstico valiosa acerca del desempeño del alumno si se registra cada opción. Si se escanean los datos, se debe asegurar que todas las respuestas queden registradas. Si se emplea un esquema uniforme se reduce el costo del ingreso manual de ítems de opción múltiple.

Los ítems de respuesta cerrada se deben calificar manualmente ya que varias respuestas diferentes pueden ser correctas. Las guías de calificación deben especificar el rango de respuestas aceptables y no aceptables. En general, el rango de opciones correctas posibles para este tipo de ítems es limitado. Dichas guías son relativamente simples y directas y el tiempo de capacitación de correctores es en general breve, pero se deben implementar y mantener controles de calidad.

Los ítems de respuesta corta abierta se califican manualmente. Como la cantidad y el rango de respuestas aceptables y no aceptables pueden ser amplios, las guías de calificación pueden llegar a ser bastante complejas y demandar mayor capacitación del corrector. El cotejo continuo durante la corrección es fundamental para controlar la calidad.

Los ítems de respuesta amplia tienen guías de calificación complejas y demandan una capacitación amplia del corrector. Para el control de calidad de estos ítems es fundamental la doble corrección continua de todos o algunos de los textos. Las guías de calificación también demandan un proceso de elaboración exhaustivo a efectos de la capacitación y deben incluir ejemplos de respuestas que se correspondan con cada uno de los niveles de la guía.

Se pueden combinar distintos formatos de ítems en la misma prueba. Por ejemplo, una prueba puede contener ítems de opción múltiple, algunos de respuesta cerrada y otros de respuesta corta abierta.

Los redactores de ítems deben tratar de asegurar que el formato de los ítems de cada prueba sea una manera adecuada y eficaz de evaluar la comprensión de un logro de aprendizaje en particular. También deben tratar de minimizar el volumen innecesario de lectura, redacción o cálculo requerido para responder un ítem en particular.

La decisión acerca del o los formatos de los ítems que se utilizarán y su proporción en una prueba debe tomarse teniendo en cuenta si el

formato es apropiado para medir un constructo o área de conocimiento o habilidad, y también considerando las restricciones prácticas, tales como la pericia necesaria para desarrollar distintos formatos de ítems y el costo de calificarlos manualmente. En la tabla 2.4 se sintetizan algunas ventajas y limitaciones de los formatos de ítems.

TABLA 2.4
Ventajas y limitaciones de los formatos de los ítems

Formato del ítem	Ventajas	Limitaciones
Opción múltiple	<ul style="list-style-type: none"> • Muchos ítems de una misma prueba pueden abordar una diversidad de logros. • Los ítems permiten establecer diferencias sutiles en el conocimiento y comprensión de los alumnos. • No se requiere calificación manual, por lo que la evaluación es relativamente económica. 	<ul style="list-style-type: none"> • Se necesita experiencia para desarrollar ítems de buena calidad. • Los alumnos no pueden demostrar su comprensión. • Los alumnos pueden adivinar la respuesta.
Respuesta cerrada (respuesta de una o dos palabras)	<ul style="list-style-type: none"> • Los alumnos generan o recuerdan la información. • La calificación manual es relativamente clara y directa. 	<ul style="list-style-type: none"> • Los ítems suelen abordar un número limitado de logros (principalmente, recuerdo y recuperación de información).
Respuesta corta abierta (respuesta de una o dos oraciones)	<ul style="list-style-type: none"> • Es posible pedir a los alumnos que demuestren niveles de comprensión altos. • Los ítems pueden abordar una diversidad de resultados. • A veces puede medir la comprensión parcial. 	<ul style="list-style-type: none"> • Se necesita experiencia para redactar claramente ítems específicos. • Se necesitan correctores capacitados y medidas de control de calidad, con lo cual aumentan los costos. • Los ítems que llevan más tiempo para que el alumno responda reducen la diversidad de logros que se pueden abordar.
Ensayo o respuesta amplia	<ul style="list-style-type: none"> • Los alumnos pueden demostrar una comprensión profunda. • Puede medir una diversidad de comprensión parcial. 	<ul style="list-style-type: none"> • Puede abordar un número limitado de logros. • Se necesitan correctores capacitados y medidas de control de calidad, lo que genera mayores costos.

TABLA 2.5
Prueba de matemáticas de Papúa Nueva Guinea - Formatos de ítems

Grado - Nivel	Opción múltiple	Respuesta corta abierta
3.º grado	20	5
5.º grado	25	10
8.º grado	25	15

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

Las pruebas de matemáticas de Papúa Nueva Guinea cambian la proporción de ítems de opción múltiple por preguntas de respuesta corta abierta, según el grado que se evalúa (véase la tabla 2.5). La prueba de 8.º grado, por ejemplo, tiene más ítems que la de 3.º grado. Este aumento se debe, en gran parte, a una mayor presencia de ítems de respuesta corta abierta.

El tiempo asignado para que los alumnos realicen la prueba debe ser suficiente para permitir que la mayoría pueda responder la mayor cantidad de ítems. Es importante recopilar información sobre el tiempo que les demanda hacer la prueba previa o los ítems de la prueba de campo. La extensión de la prueba puede variar pero los alumnos deberían poder completar la gran mayoría de los ítems aproximadamente en 40 minutos. Es posible que las pruebas con ítems en su mayor parte de opción múltiple tengan más ítems que las que tienen mayoría de respuestas cortas. Las pruebas para alumnos de escuelas secundarias pueden incluir más ítems y se les puede dar más tiempo para realizarlas. Si no están familiarizados con los formatos de los ítems en una evaluación, seguramente necesitarán más tiempo para responder.

POBLACIÓN ESCOLAR A EVALUAR

El documento marco de la prueba debe definir la población objetivo de la evaluación (por ejemplo, 4.º grado) e indicar por qué se seleccionó esa población en particular. En un país determinado, el documento marco podría justificar, por ejemplo, que se seleccione 3.º grado con el argumento de que el abandono escolar es importante después de ese grado; podría justificar que se tome la prueba en 4.º

grado porque la mayoría de los alumnos ya debería poder leer en ese grado, o podría justificar que se tome en el último grado de la escuela primaria para evaluar los logros del aprendizaje de los alumnos en este punto importante del sistema educativo. El documento marco también podría especificar subpoblaciones de alumnos que se podrían excluir de la muestra nacional, por ejemplo alumnos con necesidades educativas especiales o alumnos de escuelas pequeñas en zonas alejadas.

INFORME DE RESULTADOS

Desde el principio, se debe lograr un acuerdo con el comité director sobre cómo se deben informar los resultados. La Evaluación Nacional de Lectura en Lengua Inglesa de Irlanda informó puntajes distintos para tipos de texto y para proceso cognitivo. El documento marco de esta evaluación nacional proponía evaluar dos escalas de tipos de texto (literario e informativo) y dos escalas de proceso (recuperación e inferencia-interpretación) para 1.º grado. Para el nivel de 5.º grado, el documento proponía evaluar tres subescalas de tipos de textos (literario, informativo continuo e informativo discontinuo) y tres escalas de proceso (recuperación, inferencia e interpretación-evaluación) (Eivers et al. 2005). En la Encuesta Internacional sobre Alfabetización de Adultos se utilizaron textos discontinuos en la escala documentos para evaluar el desempeño. El marco del Estudio Internacional de Progreso en Comprensión Lectora (PIRLS) especificaba que los ítems de las pruebas se ordenarían en escalas para dos objetivos de lectura: como experiencia literaria y para obtener y aplicar la información. También proponía combinar ambas escalas y asignar un puntaje general de comprensión lectora (Campbell et al. 2001).

Debe hacerse saber al comité director que el informe de resultados por subescalas depende del análisis de los ítems. Las evaluaciones a nivel internacional y muchas a nivel nacional emplean la modelización de respuesta a los ítems para determinar si concuerdan debidamente con las subescalas. En esta etapa, podría asesorarse al comité director, a través de una introducción no técnica al concepto de

informe de resultados por niveles de rendimiento, que comúnmente se conocen como *niveles de competencia*, y se podría solicitar la opinión de sus miembros con respecto a la forma en que prefieren que se informen esos resultados. El desempeño de los alumnos se puede describir en porcentajes de ítems respondidos correctamente o de niveles, por ejemplo, *avanzado* (supera el nivel esperado) *competente* (alcanza el nivel esperado), *básico* (no alcanza el nivel esperado) o *inferior al básico* (desempeño por debajo del nivel básico). Probablemente, debe reconsiderarse la cantidad de niveles de competencia luego del análisis de las pruebas previa y final. El volumen 4 de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*, abarca la teoría de respuesta al ítem y los niveles de competencia.

El marco también debe indicar los tipos de informes de evaluación nacional que se van a publicar al finalizar dicha evaluación. Estos informes podrían incluir un informe técnico, una serie de informes resumidos para determinado público, por ejemplo, capacitadores docentes, encargados del plan de estudios y responsables de las políticas, y comunicados e informes de prensa.

CONTEXTOS

Muchos responsables de las políticas educativas emplean las evaluaciones nacionales para recopilar información contextual adicional sobre factores que pueden afectar o influir directamente en la calidad del aprendizaje escolar en áreas específicas del currículo. El comité director deberá indicar pautas generales en la selección de las variables contextuales que se pretenden evaluar. Esta información, a su vez, puede ser utilizada por la agencia implementadora como guía en el desarrollo de los cuestionarios. La información contextual puede ser particularmente útil para los responsables de las políticas si pretenden comprender las razones de las diferencias en los niveles de logro de los alumnos.

Muchas evaluaciones enfatizan los contextos de la escuela y del hogar. Los factores del hogar generalmente incluyen la condición socioeconómica, muchas veces medida según los bienes que posee el

hogar, el nivel de educación de los padres, el idioma que se habla en el hogar, la estructura y tamaño de la familia, la orientación académica a su alcance, los procesos del hogar (tales como leer a los niños u otras formas de orientación que fomenten el aprendizaje), y las relaciones hogar-escuela.² Los contextos escolares frecuentemente incluyen los recursos escolares y del aula, la administración y organización escolar, la naturaleza y nivel de capacitación de los docentes, las estrategias de enseñanza y el ambiente del aula. Algunas evaluaciones recogen datos sobre las actitudes de los alumnos respecto de la escuela y áreas temáticas individuales, los intereses y conductas de los alumnos (por ejemplo, la cantidad de tiempo dedicado a ayudar en el hogar, a trabajar o a leer por placer). Los detalles de cómo diseñar y redactar los ítems de los cuestionarios se tratan más adelante en este libro, y el CD que acompaña esta obra contiene ejemplos de ítems de cuestionarios diseñados para obtener información contextual de los alumnos, los padres, los docentes y los directivos de la escuela.

NOTAS

1. Después de su tercera implementación, esta serie de estudios se denominó Estudio Internacional de Tendencias en Matemáticas y Ciencias y conservó la sigla TIMSS.
2. Téngase en cuenta que en algunos países existe una aversión a recopilar datos de antecedentes socioeconómicos.

REDACCIÓN DE LOS ÍTEMS



Este capítulo describe las características de un buen ítem en una prueba, las pautas para redactar ítems, la estructura y la organización de los ítems para elaborar una prueba y la puntuación de los ítems.¹ También se describen las funciones del personal que participa en el desarrollo de la prueba: el equipo de redacción de ítems y otros revisores, que trabajan bajo la dirección del encargado del desarrollo de la prueba.

Se debe tener en cuenta que la calidad de una prueba depende en gran parte de la claridad con la que la prueba cumple con su propósito y la precisión con la que los ítems coinciden con unas especificaciones técnicas bien trazadas. Los ítems buenos son claros, relevantes para el currículo y se enfocan en un aspecto de la enseñanza. Establecen tareas interesantes y genuinas que son justas para los estudiantes de diferentes idiomas y con diferentes características culturales.

Un buen ítem tiene las siguientes características:

- Aborda un área clave de aprendizaje.
- Es una tarea constructiva y significativa.
- Se puede identificar con características importantes establecidas en el marco o en los documentos de diseño del proyecto.
- Es justo.

- El estímulo sigue temas centrales, no detalles periféricos o triviales.
- Le dice claramente a los estudiantes lo que deben hacer.
- Es autosuficiente y no depende del entendimiento de un ítem previo.
- Si es sobre vocabulario, se dirige al significado de la palabra en su contexto, en lugar del conocimiento general.
- Preferentemente, se expresa en términos positivos; los negativos tienden a provocar confusión.

Los redactores de ítems pueden beneficiarse enormemente del examen de modelos de ítems de alta calidad relevantes. Muchas organizaciones examinadoras publican ítems de muestra en Internet. Los ítems de prueba publicados también pueden provenir de pruebas internacionales, tales como el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS), el Programa para la Evaluación Internacional de Alumnos (PISA) y el Estudio sobre el Progreso Internacional de la Competencia en Lectura (PIRLS), así como evaluaciones internacionales de otros países, como la Evaluación Nacional del Progreso Educativo (NAEP) de EE. UU. El disco compacto (CD) complementario incluye muchos ejemplos de ítems de estas y otras fuentes. En el apéndice B se detallan los sitios web. Los ítems de prueba publicados se pueden utilizar en otras pruebas, siempre que el material contenido y la redacción del ítem sean apropiados. Este material puede ser una fuente económica y útil de ítems de prueba. Estos ítems (junto con las respuestas) suelen encontrarse en Internet. Por lo tanto, tales ítems no deben utilizarse si es probable que los estudiantes tengan acceso a ellos antes de realizar la prueba de evaluación nacional. También se puede obtener permiso de las autoridades competentes para utilizar ítems apropiados de pruebas no divulgados. Este enfoque puede ser más económico que el desarrollo de ítems. Sin embargo, los expertos en currículo deberán revisar tales ítems y examinarlos previamente para determinar si son adecuados.

El desarrollo de la pericia en la redacción de ítems para cada sección del diseño de un proyecto lleva tiempo. Los redactores de ítems deben tener un entendimiento común de la terminología de los ítems y de lo que estos deben medir. Para alcanzar este entendimiento,

deben tratar de clasificar cada borrador de ítem mientras lo elaboran, mediante los siguientes criterios:

- Formato del ítem (por ejemplo, de opción múltiple, de respuesta cerrada o de respuesta abierta).
- Tipo de texto (para una prueba de comprensión lectora) (por ejemplo, narrativo o expositivo).
- Nivel de grado (por ejemplo, 5.º grado).
- Resultado del aprendizaje (por ejemplo, suma de números enteros de dos dígitos o identificación de la idea principal en una historia).
- Proceso cognitivo (por ejemplo, conocimiento, memoria, interpretación o síntesis).

No siempre se puede saber el nivel del proceso cognitivo relacionado con la respuesta a un ítem. Si los estudiantes no estuvieran familiarizados con un proceso, tal como el resumen de un párrafo informativo, tal vez les sea necesario un nivel de procesamiento superior al que se necesitaría si estuvieran acostumbrados a hacer resúmenes.

DIFICULTAD DE LOS ÍTEMS

Obtener un nivel de dificultad apropiado para los ítems es una tarea exigente para la mayoría de los redactores. En muchos países, el contenido del currículo previsto es demasiado difícil para la mayoría de los estudiantes. Por consiguiente, los redactores de ítems a menudo deben redactar un borrador de varios ítems para medir habilidades más fáciles que aquellas enumeradas en los currículos previstos. Por ejemplo, las pruebas de rendimiento académico en matemáticas diseñadas para 5.º grado con frecuencia incluyen ítems basados en objetivos que los estudiantes deberían haber alcanzado en 3.º y 4.º grado.

Es más probable que los docentes con experiencia tengan mejor percepción sobre los probables niveles de dificultad de los ítems que los funcionarios educativos o los académicos. Sin embargo, aunque el juicio del docente es útil, no es adecuado. Es esencial someter a prueba los ítems previamente en muestras de estudiantes más o menos

similares a aquellos existentes en la población objetivo para obtener datos iniciales objetivos sobre los niveles de dificultad de los ítems. Esto puede ayudar a evitar el error frecuente de realizar pruebas con ítems que terminan siendo demasiado difíciles.

Muchos factores pueden ocasionar que ciertos ítems, que el redactor consideró simples, resulten bastante difíciles. Asimismo, algunos ítems que se diseñaron para ser difíciles pueden redactarse o presentarse de tal modo que pueden resultar bastante fáciles. En el caso de los ítems que usan el formato de opción múltiple, los redactores deben evitar lo siguiente:

- Introducir en el enunciado y la clave pistas gramaticales o lógicas que apunten a la respuesta correcta, tal como un enunciado que contenga un sustantivo en singular para opciones en plural excepto una.
- Introducir términos absolutos, tales como “siempre” o “nunca”, que pueden excluir algunas opciones o llevar a la respuesta correcta.
- Redactar la respuesta correcta de manera mucho más larga o detallada que las otras opciones.
- Incluir palabras clave o frases del material de estímulo en la opción correcta, pero no en las otras opciones.
- Disponer las opciones en un orden ilógico o con un patrón confuso.
- Dar pie a que las opciones y la clave se superpongan considerablemente, de modo que la distinción de la “mejor” respuesta dependa de las habilidades lingüísticas y no del conocimiento de lo que se evalúa.
- Incluir preguntas que se puedan responder sin hacer referencia al estímulo.

En el caso de los ítems con formato de respuesta construida, es importante desarrollar criterios de puntuación claros que obtengan lo que el ítem intenta medir.

SESGO DE LOS ÍTEMS

Los estudiantes son portadores de experiencias culturales y conocimientos diversos. No deberían ser penalizados ni favorecidos por experiencias que no incumben al conocimiento, las habilidades y el entendimiento

que la prueba intenta evaluar. Por ejemplo, los ítems sobre un deporte masculino popular pueden desfavorecer a las mujeres.

Los ítems también pueden estar sesgados si molestan a ciertos estudiantes, pero no a otros. El material de estímulo no debe herir susceptibilidades éticas, culturales o de otro tipo. No se debe dar cabida a que algunos estudiantes se ofendan, asusten o molesten debido al material de la prueba. El encargado del desarrollo de la prueba debe sensibilizar a los redactores de ítems sobre las diferentes formas de sesgo. Se debe alentar a los comités de revisión a que estén atentos a ítems de prueba o cuestionario que puedan estar sesgados o que puedan ofender. El cuarto volumen de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*, presenta una técnica estadística que se puede utilizar para ayudar a detectar ítems sesgados en las etapas de prueba previa y prueba final.

MATERIAL DE ESTÍMULO

El material de estímulo contextualiza el ítem. Puede ser un segmento de un texto, un diagrama, un gráfico, una tabla, un mapa, un cuadro o cualquier combinación de los mismos.

La mayor parte del desarrollo de una prueba comienza con la selección o la creación del material de estímulo apropiado. Por lo general, las pruebas de comprensión lectora se basan en textos de gran extensión que se prestan a una serie de ítems o una unidad que aborda una diversidad de habilidades relevantes. Las pruebas de matemáticas y ciencias pueden incluir un material de estímulo breve, tal como números para sumar o una ecuación que se debe completar. Los ítems de matemáticas y ciencias también pueden incluir un estímulo más complejo, como un gráfico, un cuadro, una tabla o un diagrama con una serie de ítems asociados que abordan una cierta serie de habilidades.

El material de estímulo debe presentar de forma clara las características principales que se deben evaluar. No debe contener detalles superfluos, repetitivos o innecesarios. Un buen material de estímulo tiene las siguientes características:

- Es sustancial y vale la pena examinarlo con atención.
- Es probable que sea de interés para la audiencia objetivo.

- Está bien escrito y bien diseñado.
- Es óptimo en cuanto a su nivel de desafío; ni demasiado difícil ni demasiado fácil.
- No presenta desafíos engañosos.
- Es correcto en cuanto a los hechos.
- Ofrece la oportunidad de usar preguntas indagatorias.
- Es autosuficiente.

Cuando sea apropiado, es importante proporcionar contexto para el material de estímulo. El contexto se puede indicar mediante un encabezado o una introducción breve. Por ejemplo, un fragmento de una novela de ciencia ficción se puede presentar de la siguiente manera: “Este fragmento pertenece a una novela que se sitúa en el futuro”.

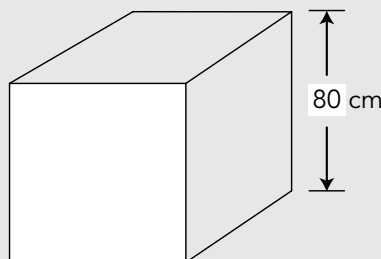
Es preferible que las imágenes sean una parte intrínseca del material de estímulo, que proporcionen un significado adicional. Si las imágenes se incluyen simplemente como decoración, no deben ayudar a los estudiantes a entender del texto.

A veces, el material de estímulo crea un contexto artificial e innecesario para un ítem. El recuadro 3.1 contiene material irrelevante. Este ítem trata en realidad sobre el área de una superficie. En la vida real, Irene no estaría preocupada por la menor cantidad de papel que usaría. En realidad, podría necesitar una pequeña cantidad adicional para doblar los bordes. Los estudiantes más aptos podrían responder a este ítem de forma incorrecta al calcular el papel adicional. El ítem se

RECUADRO 3.1

Ejemplo de material de estímulo irrelevante

Irene tiene que envolver con papel esta caja. ¿Qué cantidad mínima de papel deberá usar?



redacta mejor simplemente de la siguiente manera: “La longitud de un lado de un cubo es de 80 cm. ¿Cuál es el área de la superficie del cubo?”

Un estímulo que intenta describir el contexto de la vida real debe ser preciso en cuanto a los hechos. La información en el recuadro 3.2 probablemente sea inexacta. Por lo general, los seres humanos no presentan el tipo de patrón de crecimiento mostrado. Los niños que tienden a ser altos a menudo demuestran este atributo desde muy pequeños. Si el material de estímulo requiere patrones de crecimiento desiguales, puede ser preferible utilizar plantas en lugar de seres humanos con fines comparativos.

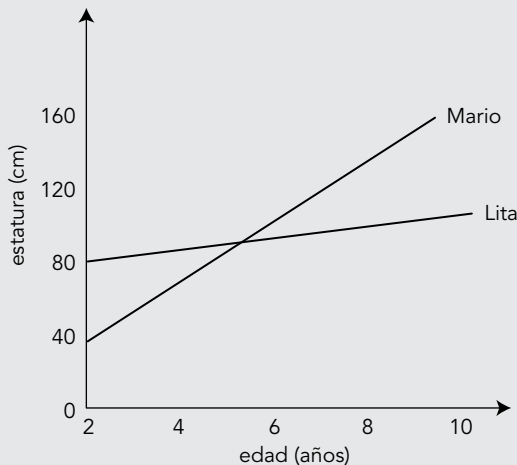
Los ítems se deben redactar con el lenguaje más simple y claro posible. La redacción debe ser simple para que sea razonable esperar que los estudiantes lo entiendan:

- Evitar vocabulario complejo.
- Evitar oraciones largas.
- No utilizar oraciones complicadas.
- No utilizar una lógica difícil.

RECUADRO 3.2

Ejemplo de un ítem con información inexacta o engañosa

El gráfico muestra el cambio en la altura de Mario y Lita a medida que crecen.



- Evitar las preguntas ambiguas o vagas.
- Evitar los dobles negativos.
- Evitar las incoherencias (por ejemplo, el uso de diferentes unidades de medición en las opciones o diferentes términos para referirse a lo mismo).
- No utilizar expresiones vagas o términos extraños que no se definan de forma adecuada.

FORMATO DE LOS ÍTEMS

Se describen dos formatos importantes: de opción múltiple y de respuesta corta (consulte el capítulo 2).² Considere utilizar un formato de opción múltiple a los siguientes efectos:

- Limitar la cantidad de opciones.
- Obtener una respuesta concisa.
- Evitar que los estudiantes tengan que copiar grandes secciones del texto de estímulo.
- Cubrir un amplio número de temas de forma eficiente.

Considere utilizar un formato de respuesta corta (de respuesta cerrada o de respuesta abierta) a los siguientes efectos:

- Evaluar el significado que los estudiantes deben generar por su cuenta.
- Evaluar los diferentes niveles de entendimiento, mediante un ítem de crédito parcial que otorgue un puntaje completo para una respuesta que muestre la comprensión de una idea compleja y que otorgue un puntaje parcial para una respuesta que muestre la comprensión de una parte simple de la idea.
- Evaluar un número restringido y claramente definible de posibles respuestas correctas.
- Buscar una respuesta en una situación donde la respuesta correcta aparecería con claridad en un formato de opción múltiple debido a la falta de opciones incorrectas convincentes.

Para responder ítems de respuesta corta, los estudiantes deben tener un dominio adecuado del vocabulario y la expresión. No utilice

preguntas de respuesta corta si es probable que los estudiantes copien una gran parte del texto de estímulo.

Redacción de los ítems de opción múltiple

Un ítem de opción múltiple consta de un enunciado y cierta cantidad de opciones de respuesta. A veces, cuando es necesaria una respuesta verdadero/falso, solo se requieren dos opciones. Sin embargo, estos ítems son algo ineficaces. Es más común proporcionar cuatro o cinco opciones. La opción correcta es la *clave* y las opciones incorrectas son los *distractores* (recuadro 3.3).

El *enunciado* de un ítem puede adoptar varias formas, tales como:

- Una oración inconclusa.
- Una pregunta definida de forma explícita.
- Una oración a la que le falta información (Carlos tiene _____ hermanos).

Si el enunciado es una oración incompleta, debe contener suficiente información como para indicar la naturaleza de la pregunta. El estudiante no debe necesitar leer los distractores para inferir la pregunta. Todas las opciones de una oración incompleta deben:

- Ser gramaticalmente coherentes con el enunciado.
- Estar escritas en un estilo similar.
- Estar puntuadas correctamente.
- Comenzar con una letra en minúscula y terminar con un punto.

RECUADRO 3.3

Ejemplo de un ítem de opción múltiple

Tula tiene 3 flores. Papá le da 2 flores más. ← enunciado

¿Cuántas flores tiene Tula en total?

- A 2 ←
B 3 ←
C 4 ←
- distractores

D 5 ← clave (respuesta correcta)

Al redactar ítems de opción múltiple se deben recordar los siguientes puntos:

- *Puntuar las oraciones completas de manera correcta.* En el recuadro 3.4, todas las opciones son oraciones completas con la puntuación apropiada.
- *Puntuar las listas de manera apropiada.* En el recuadro 3.5, las opciones son listas de palabras. Estas opciones no tienen signos de puntuación.
- *Minimizar la cantidad de lectura.* Para minimizar la cantidad de lectura requerida, el redactor del ítem debe colocar la mayor parte posible del ítem en el enunciado (consulte el recuadro 3.6).
- *Evitar enunciados negativos.* Debido a que un enunciado negativo causa confusión, se debe evitar su uso. Si el enunciado solo se puede

RECUADRO 3.4

Puntuación en oraciones completas

¿Qué opinó Miho del mercado?

- A Estaba lleno de gente y la comida era buena.
- B Era económico y la comida era deliciosa.
- C La comida era buena, pero no había nadie allí.
- D Era económico, pero la comida no era muy buena.

RECUADRO 3.5

Puntuación en una lista

¿Durante cuánto tiempo se quedará Joe en la casa de su abuelo en Bali?

- A una semana
- B dos semanas
- C un mes
- D dos meses

RECUADRO 3.6**Minimizar la cantidad de texto que debe leerse**

¿Durante cuánto tiempo se quedará Joe en la casa de su abuelo en Bali?

No esto

Sino esto

Carl fue

Carl y su familia fueron

A al río con su familia.

A al río.

B a la playa con su familia.

B a la playa.

C al campo con su familia.

C al campo.

D a las montañas con su familia.

D a las montañas.

RECUADRO 3.7**Ítem con un enunciado negativo**

¿Qué le dijeron los padres de Mario a Mario que **no** podía tener en su casa?

A un perro

B sus zapatos malolientes

C la manta del caballo

D una canasta con frutas

expresar de forma negativa, resalte la palabra “no” con negrita o cursiva (consulte el recuadro 3.7). Si el enunciado negativo es inevitable, las opciones nunca deben ser negativas.

- *Variar el uso de distractores emparejados.* Los métodos de construcción de distractores deben variar a lo largo de la prueba para que los patrones no salten a la vista y ayuden al estudiante. Por ejemplo, no es aconsejable emparejar la clave (B) con su opuesto (A) (recuadro 3.8). Si el patrón en el recuadro 3.8 se repite durante la prueba, se hará evidente para algunos estudiantes familiarizados con las pruebas que solamente necesitan tener en cuenta los distractores emparejados (A y B). Una solución es redactar algunos ítems en los que la clave no sea uno de los opuestos emparejados. Otra solución es incluir dos pares de opuestos en un ítem, como se muestra en el recuadro 3.9.

RECUADRO 3.8**Distractores mal emparejados**

A Tom no le gustó la chaqueta porque

- A era demasiado grande.
- B era demasiado pequeña.
- C era del color equivocado.
- D no era lo suficientemente abrigada.

RECUADRO 3.9**Uso de pares en los distractores**

A Tom no le gustó la chaqueta porque era demasiado

- A grande.
- B liviana.
- C pequeña.
- D pesada.

- *Evitar el uso de determinados distractores.* Los distractores que contienen palabras como *siempre* y *nunca*, *ninguna de las anteriores* y *todas las anteriores* deben evitarse porque a menudo es fácil que el estudiante las descarte.
- *Utilizar una cantidad de distractores adecuada.* Se deben desarrollar ítems con una clave y cuatro distractores convincentes (cinco opciones en total), si es posible, y luego, examinar previamente todos los distractores. Se deben usar los distractores que tienen las mejores propiedades estadísticas (consúltese el volumen 4 de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*).
- *Variar la posición de la clave.* La posición de la clave debe variar de un ítem al otro. No debe haber un patrón obvio en cuanto a su ubicación. Las opciones se pueden organizar de la más corta a la más

larga, de la más larga a la más corta o de modo aleatorio. Es necesario asegurarse de que la clave no sea siempre la opción más larga.

Las buenas opciones tienen las siguientes características:

- Son de extensión similar y están escritas con un estilo similar a la clave. La clave no debe destacarse por su extensión, redacción u otra cualidad superficial en relación con los distractores.
- Varían en estilo de ítem a ítem. No son repetitivas.
- No dan pistas de la respuesta a otro ítem.
- No incluyen distractores parcialmente correctos, tales como opciones emparejadas, donde cada distractor contiene una opción incorrecta y una correcta.
- No desorientan ni confunden mediante la falta de claridad o ambigüedad.
- No se superponen en cuanto al significado. Los distractores deben tener significados diferentes entre sí. Los distractores no deben ser sinónimos. Un significado particular en un distractor no debe estar incluido en el significado general de otro distractor.
- Incluyen una clave que es indiscutiblemente correcta o una respuesta precisa y justificable a la pregunta, y no simplemente la mejor de las opciones que se presentan.
- Tienen distractores que son indiscutiblemente incorrectos, aunque razonables y convincentes. Todo distractor que sea ridículamente erróneo reduce el número de opciones reales disponibles para el estudiante y no contribuye en nada al ítem.

Redacción de los ítems de respuesta corta

Los ítems de respuesta corta deben enfocarse claramente en dejar patente la habilidad que se desea evaluar. Los ítems de respuesta corta son claros y precisos. Las guías de corrección se deben desarrollar al mismo tiempo que los ítems. El CD complementario incluye ejemplos de guías de corrección para ítems de respuesta corta. Por ejemplo, consúltese *PISA Mathematics Released Items 2006* [Ítems publicados de matemáticas PISA 2006] y *Reading International Grade 4 PIRLS 2001 Sample Items Scoring Guide* [Guía de corrección de ítems de muestra de comprensión lectora internacional de 4.º grado PIRLS 2001].

En general, los ítems de respuesta corta se clasifican como de respuesta abierta cuando una respuesta correcta requiere una o dos oraciones, o varios agregados a un diagrama. Los ítems de respuesta abierta a menudo tienen un número de respuestas correctas posibles. Los ítems de respuesta corta se clasifican como de respuesta cerrada cuando una o dos palabras, o una línea en un diagrama, son suficientes para la respuesta. En general, los ítems de respuesta cerrada tienen un número muy limitado de respuestas correctas.

Los ítems de respuesta abierta deben abordar habilidades sustanciales en áreas clave del currículo para justificar el tiempo de examen que los estudiantes necesitarán para responderlos. Aun así, los estudiantes deben dar respuestas correctas breves a los ítems de respuesta abierta. La mayor parte del tiempo que los estudiantes le dedican a un ítem debe ser para conseguir una solución en lugar de anotar la respuesta.

Es importante tener en cuenta qué respuesta podría ser errónea en un ítem de respuesta corta. Si fuera probable que todas las respuestas coherentes imaginables sean correctas, tal vez el ítem contribuya poco a la evaluación de una habilidad en particular. El ítem se debe estructurar para garantizar que haya respuestas incorrectas convincentes.

Asegúrese de que los ítems de respuesta corta tengan más de dos respuestas posibles. Los ítems para los que solo existen dos opciones posibles, tales como “abierto” o “cerrado”, le dan a los estudiantes un 50 % de posibilidades de adivinar la respuesta correcta. Este tipo de ítem se puede extender al pedirles a los estudiantes que justifiquen sus respuestas. Al ítem, entonces, podría asignársele un puntaje de acuerdo con la selección correcta de “abierto” o “cerrado” y la explicación. Los estudiantes que seleccionan la opción correcta sin proporcionar una explicación recibirían un puntaje de cero.

Los ítems no deben proporcionar una ayuda exhaustiva al lector en la comprensión del significado del estímulo. Por ejemplo, un ítem no debe resumir las ideas clave de un párrafo del estímulo ni explicitar una inferencia en el estímulo. Es preferible citar el estímulo, en lugar de resumir o interpretar su significado.

Un peligro que existe con los ítems de respuesta abierta es que los estudiantes los respondan superficialmente. Por ejemplo, la respuesta “porque es importante” podría ser una respuesta técnicamente correcta, pero insuficiente para varias preguntas. Algunas veces, puede incluirse

una respuesta potencialmente superficial a una pregunta para eliminarla del número de respuestas correctas posibles. Por ejemplo, un ítem se puede redactar de la siguiente forma: “¿En esta historia, por qué es importante el accidente de Jemina?” Los estudiantes no pueden responder a este ítem con “porque el accidente es importante”. Con frecuencia, instrucciones tales como “explique su respuesta” o “justifique su respuesta” son necesarias en los ítems de respuesta abierta para evitar una respuesta corta de “sí” o “no”.

Un ítem de respuesta corta eficaz debe establecer una tarea clara y específica que busque una respuesta específica. El ítem les debe permitir a los estudiantes que demuestren su dominio de la habilidad requerida de un modo razonablemente rápido. El ejemplo en el recuadro 3.10 no logra alcanzar ninguno de estos objetivos. No se le dice a los estudiantes

RECUADRO 3.10

Ítem de respuesta abierta confuso con indicaciones poco claras

Se debe hacer una caja abierta a partir de un papel rectangular duro, de 150 cm por 100 cm, cortando cuadrados del mismo tamaño en cada esquina y utilizando cinta adhesiva para unir los bordes. ¿Qué tamaño de cuadrado se cortará de la esquina? Justifique el(los) motivo(s) para elegir este tamaño.

Guía de corrección: El número de puntos varía de 0 a 3.

3 puntos: describe un cuadrado con lados de 20 cm y una caja con dimensiones de 110 cm por 60 cm por 20 cm; también explica que este tamaño de caja tiene la capacidad más grande.

2 puntos: describe que en cada esquina debe cortarse un cuadrado de 20 cm, pero no ofrece explicación.

1 punto: describe todo tamaño de cuadrado posible con un lado inferior a 50 cm.

0 puntos: ofrece dimensiones del cuadrado superiores a 50 cm (una respuesta imposible).

9 faltantes.

RECUADRO 3.11**Un buen ejemplo de un ítem con respuesta cerrada**

Todas las personas cavan a la misma velocidad.

Una persona puede terminar de cavar un jardín en 12 horas.

Dos personas pueden terminar de cavar el mismo jardín en 6 horas. ¿Cuánto tiempo les lleva a 4 personas? _____

¿Cuánto tiempo le lleva a x personas? _____

que necesitan hacer una caja lo más grande posible. Sin embargo, esta respuesta es el criterio para un puntaje de 3 puntos. La calidad del ítem también es deficiente porque consume demasiado tiempo. Las habilidades que se evalúan no justifican la cantidad de tiempo que los estudiantes necesitarían para experimentar con las posibilidades y llegar a la respuesta correcta. El problema necesita simplificarse para que los estudiantes puedan demostrar las habilidades relevantes de forma eficiente.

Cuando un estudiante no responde a un ítem en particular por una de varias razones (tal como omitirlo porque se consideró difícil o por no tener la oportunidad de intentarlo porque el ítem no estaba en el cuadernillo de prueba asignado), con frecuencia se asigna un valor de código (no un puntaje) de 9 para denotar los datos faltantes. Los datos faltantes se explican con más detalle en el volumen 3, *Implementación de una evaluación nacional del rendimiento académico*.

Los ítems de respuesta corta deben redactarse con claridad y sencillez (consulte el recuadro 3.11).

Desarrollo de las guías de corrección para los ítems de crédito parcial

Las respuestas a algunas preguntas de respuesta corta tienen dos o más categorías de respuesta correcta. Se conocen como ítems de crédito parcial. La guía de corrección debe distinguir entre las respuestas más exhaustivas, precisas o sofisticadas y las respuestas incompletas o parcialmente correctas. A las mejores respuestas se les asigna un puntaje más alto. El ejemplo en el recuadro 3.12 es la guía de corrección para

un ítem de crédito parcial que consiste en dibujar un cuadrado y que puede recibir un puntaje de hasta 3 puntos.

Los siguientes tipos de ítems pueden recibir puntaje como ítems de crédito parcial:

- Se les pide a los estudiantes que detallen dos motivos para el comportamiento de un personaje. Los estudiantes que dan dos motivos correctos reciben un puntaje de 2 y aquellos que dan un motivo correcto reciben un puntaje de 1.
- Se les da un puntaje más alto a los estudiantes que demuestran un entendimiento más sofisticado; por ejemplo, un puntaje de 2 puntos en una prueba de comprensión lectora podría reflejar la comprensión de la ironía en un fragmento de texto, mientras que se asigna un puntaje de 1 punto a una comprensión lectora literal.
- Un puntaje de 2 puntos puede incluir la identificación tanto de la causa como de la consecuencia, mientras que un puntaje de 1 punto requiere la identificación correcta de solamente una de ellas.
- En matemáticas, se asigna un puntaje de 3 a la solución correcta a un problema y una explicación apropiada del método de

RECUADRO 3.12

Ítem de crédito parcial

La longitud del lado de un cuadrado es de 10 cm.

Dibuje este cuadrado en el espacio que aparece a continuación. Use una regla.

Guía de corrección:

3 puntos: dibuja un cuadrado con 4 lados de 10 cm de longitud y 4 ángulos rectos.

2 puntos: dibuja un rectángulo con 2 lados de 10 cm de longitud y 4 ángulos rectos.

1 punto: dibuja una figura de 4 lados con 2 lados de 10 cm de longitud, pero sin ángulos rectos.

0 puntos: dibuja cualquier otra figura.

9 faltantes.

resolución; un puntaje de 2 a la solución correcta sin una explicación y un puntaje de 1 a una descripción de un método apropiado con cálculos incorrectos.

La distinción entre los puntajes de crédito total o parcial debe ser clara. Asegúrese de que los ejemplos de las respuestas de 1 punto, proporcionados en la guía de corrección, no estén mal redactados ni sean respuestas breves que en realidad satisfacen los criterios de 2 o 3 puntos. También es importante aclarar la diferencia entre las respuestas de 1 punto y las respuestas incorrectas. Esta distinción puede ser la más difícil de hacer al asignarle un puntaje a algunos ítems de crédito parcial.

Los siguientes ejemplos muestran que los ítems de respuesta abierta de respuesta corta no siempre permiten un crédito parcial. El ítem en el recuadro 3.13 muestra que aunque los estudiantes pueden brindar una variedad de respuestas, pueden obtener un puntaje de 1 o de 0 puntos.

RECUADRO 3.13

Ejemplo de un ítem de respuesta abierta con guía de corrección

John y Michael encuentran un árbol con 400 mangos.

John dice que ahora existe una probabilidad del 160 % de que Michael le acierte a un mango. ¿Está de acuerdo con John o no?

Explique.

Guía de corrección:

1 punto: *No está de acuerdo Y se refiere al límite del porcentaje.*

No está de acuerdo porque no se puede obtener el 160 %.

No está de acuerdo porque es imposible.

No está de acuerdo porque el 100 % es lo máximo que se puede obtener.

0 puntos: *Está de acuerdo (con o sin explicación).*

No está de acuerdo Y no se refiere al límite del porcentaje.

No está de acuerdo porque hay más de 160 mangos.

9 faltantes.

Fuente: Departamento de Educación de Filipinas 2004.

Es importante examinar previamente los ítems de crédito parcial para garantizar que las categorías de crédito parcial sean sólidas desde el punto de vista estadístico (consúltese el capítulo 5). En el recuadro 3.14 se presenta un ejemplo de ítem de respuesta corta con una guía de corrección colapsada. En la prueba previa, se asignó un puntaje de 2 puntos para las respuestas de “80 %” u “80 por ciento” y de 1 punto para los estudiantes que simplemente escribieron “80”. Las estadísticas muestran que los estudiantes que dieron una respuesta de “2 puntos” tuvieron un puntaje promedio mucho más elevado en la prueba de matemáticas y los estudiantes que dieron una respuesta de “1 punto” tuvieron un puntaje promedio general similar al de quienes obtuvieron un puntaje de cero en este ítem. Como resultado de esta información obtenida de la prueba previa, se cambió la guía de corrección y se otorgó 1 punto a los estudiantes que respondieron con “80 %” u “80 por ciento” y cero puntos a los estudiantes que respondieron “80” o dieron otra respuesta inaceptable.

RECUADRO 3.14

Ejemplo de un ítem de respuesta cerrada con guía de corrección

Árbol de mango

Michael está tratando de acertarle con su tirachinas a algunos mangos en su granja. Cuando el árbol tiene 50 mangos, existe una probabilidad del 20 % de que le acierte a uno.

Su probabilidad de darle a un mango se duplica cuando el número de mangos se duplica.

Calcule la chance de que Michael le dé a un mango en un árbol con 200 mangos.

Guía de corrección:

1 punto: 80 % u 80 por ciento.

0 puntos: cualquier otra respuesta, incluso 80 sin especificar “%” o “por ciento”.

9 faltantes.

Fuente: Departamento de Educación de Filipinas 2004.

Redacción de ítems por unidades

Las unidades son grupos de ítems con un estímulo en común. Una unidad puede constar de una historia breve o un gráfico seguido de un conjunto de preguntas. A la redacción de ítems asociados con unidades se aplican los mismos principios básicos para la redacción de ítems de opción múltiple o de respuesta corta.

Al redactar el borrador de un ítem basado en unidades se deben tener en mente ciertos puntos:

- Los ítems deben ser independientes el uno del otro. Los estudiantes no deben necesitar responder un ítem de forma correcta para responder otros ítems correctamente.
- Los ítems no se deben superponer. Cada ítem debe evaluar claramente un aspecto diferente del estímulo.
- Los ítems deben evaluar un número de habilidades. Por ejemplo, los ítems no deben evaluar de forma repetida la recuperación de información directamente expresada o la idea principal de cada párrafo del estímulo.
- Los ítems de una unidad deben cubrir un rango de niveles de dificultad y comenzar, por lo general, con un ítem fácil.
- La información proporcionada en el enunciado o en las opciones múltiples de un ítem no deben ayudar al estudiante a que responda otro ítem.
- Los ítems deben evaluar aspectos sustanciales (y evitar aspectos triviales) del estímulo.
- Los ítems deben estar en la misma página que la unidad o en la página contigua (en el caso de estímulos extensos).

Las unidades con ocho o más ítems suelen tener ítems duplicados, solapados o triviales. Algunos ítems se pueden eliminar durante su presentación ante los comités encargados de revisarlos. Otra posibilidad es que el equipo de desarrollo de la prueba elabore dos formularios para la prueba previa, con la mitad de los ítems en un formulario y el resto en el segundo.

La sección de lengua en el CD complementario contiene muchos ejemplos de unidades seguidas de un conjunto de preguntas. (Consúltese, por ejemplo, los ítems del CD luego de “Hare Heralds the Earthquake” [*Hare predice el terremoto*] en *Reading International*

Grade 4 PIRLS 2001 Sample Items [Ítems de muestra de comprensión lectora internacional de 4.º grado PIRLS 2001] o “Petra’s Deal” [El acuerdo de Petra] en Reading Australia Year 3 Sample Questions [Preguntas de muestra de comprensión lectora de 3.º año de Australia]].

ÍTEMS DE PRÁCTICA

Los ítems de práctica son esenciales para garantizar que los estudiantes no sean penalizados por su falta de familiaridad con el formato de los ítems o con el modo en que deben presentar sus respuestas a las preguntas de la prueba. En general, el examinador revisa los ítems de práctica junto a los estudiantes, de acuerdo con instrucciones muy específicas incluidas en el manual de administración de las pruebas.

Cuando las pruebas incluyen ítems de respuesta corta, es especialmente importante que los estudiantes comprendan qué tipo de respuesta se espera que den. Por ejemplo, los estudiantes necesitan entender dónde deben escribir sus respuestas para completar una secuencia numérica o cuánto se espera que escriban como respuesta a un ítem que requiere una explicación. Se les debe decir a los estudiantes que no serán sancionados por cometer errores ortográficos o gramaticales mínimos, a menos que sea parte de lo que se está evaluando. Las instrucciones deben alentarlos a que intenten responder todos los ítems.

Se deben desarrollar preguntas de práctica para todos los formatos de respuesta de la prueba. Por este motivo, las preguntas de práctica suelen escribirse hacia el final de la fase de desarrollo de la prueba, cuando ya se conocen los tipos de preguntas que aparecerán en la prueba. Los ítems de práctica deben ser *muy fáciles*; por ejemplo, se les podría pedir a los estudiantes que escriban la respuesta a $2 + 2$ en la línea que se muestra junto a la suma:

$$2 + 2 = \underline{\hspace{2cm}}$$

El énfasis se encuentra en la manera en la que los estudiantes presentan sus respuestas. En este caso, los estudiantes deben escribir sus respuestas en la línea.

El comité encargado de revisar los ítems debe revisar todos los ítems de práctica, que también se deben examinar previamente. Si se utilizan varios formularios de prueba previa, con cada formulario se deben usar los mismos ítems de práctica.

DIAGRAMACIÓN Y DISEÑO DE LOS ÍTEMS

La diagramación y el diseño del ítem son cruciales para la claridad y el atractivo de una prueba. Es más probable que los estudiantes intenten resolver los ítems en una prueba que esté bien presentada y sea fácil de leer. Los diseñadores de la prueba deben adoptar un estilo de diagramación coherente y agradable. Se pueden usar un diseño y diagramación profesionales para crear todas las imágenes. El CD complementario proporciona muchos ejemplos de ítems bien presentados respaldados por ilustraciones de buena calidad.

Pautas básicas

A continuación, aparecen algunas pautas básicas para una diagramación y diseño de examen exitosos.

- Use un tamaño de fuente grande (por ejemplo, 36 puntos) para numerar los ítems, de modo que los estudiantes puedan ubicar cada ítem con facilidad.
- Deje un espacio adecuado si los estudiantes deben escribir una respuesta. (Esto es especialmente necesario para los estudiantes de escuela primaria, que pueden tener una caligrafía grande).
- Deje suficiente espacio entre los ítems para que los estudiantes puedan ver con claridad dónde termina un ítem y dónde comienza el siguiente.
- Utilice la longitud de la línea para la respuesta de un ítem para indicar cuánto se espera que escriban los estudiantes. Una línea corta es adecuada para una respuesta de una palabra. Dos o tres líneas más extensas sugieren que el estudiante debe escribir una o dos oraciones como respuesta al ítem.
- Dele a cada ítem una etiqueta única e imprima esta etiqueta al lado del número de ítem en una fuente pequeña en escala de grises sobre el margen izquierdo. Una etiqueta de identificación única ayudará a garantizar el seguimiento preciso de los ítems. Los números de los ítems pueden variar, en especial si los ítems aparecen en varios cuadernillos de examen.
- Sea coherente en cuanto al uso de comillas simples o dobles, cursiva, subrayado, negrita y mayúsculas.

- Asegúrese de que la diagramación y las imágenes utilizadas en los ítems de la prueba previa sean lo más cercanas posible a la presentación de los ítems en el formulario final de la prueba. Los cambios en el diseño y la diagramación de los ítems pueden afectar la dificultad de un ítem.

Calidad de las imágenes

Las imágenes de una prueba deben ser claras, con lenguaje y encabezados simples y apropiados. Un artista gráfico debe ilustrar las imágenes. Las imágenes escaneadas o las que se toman de Internet suelen no ser de calidad suficiente. En general, las fotografías son también inadecuadas porque aumentan los costos de impresión. Generalmente, un artista gráfico debe volver a dibujar las fotografías. Al utilizar el mismo artista gráfico para dibujar todas las imágenes, fotográficas o de otro tipo, se le da coherencia a la diagramación y al diseño.

Siempre que sea posible, se deben utilizar imágenes para mejorar la claridad y reducir la cantidad de palabras en un ítem (consúltese el recuadro 3.15). Las imágenes simples son las más eficaces. La imagen debe presentar el concepto de forma clara e impecable (consúltese el recuadro 3.16). No necesitan tener un aspecto real. Si es posible, se deben usar dibujos de líneas y se deben evitar los sombreados extensos.

Los gráficos y los mapas se deben etiquetar de forma clara y coherente (consúltense los recuadros 3.17 y 3.18).

Como se muestra en el recuadro 3.19, dejar espacios en los gráficos permite que el material de estímulo se lea con facilidad.

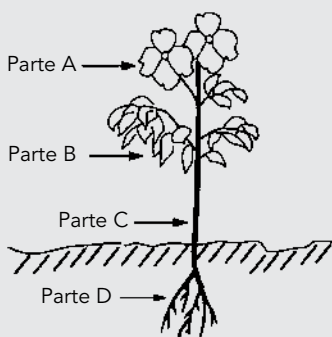
Al momento de decidir el tamaño de la fuente y la longitud de la línea, los diseñadores de la prueba deben tener en cuenta lo siguiente:

- Usar una fuente de 14 puntos para 3.º y 4.º grado, y una fuente de 12 puntos para grados superiores.
- Reducir el ancho de los textos de estímulo a aproximadamente entre 10 y 14 palabras por renglón.
- Asegurarse de que los cortes de línea se ubiquen en lugares razonables. No permitir que una única palabra quede sola en una nueva línea.

RECUADRO 3.15**Uso de ilustraciones para reducir la cantidad de palabras**

La siguiente imagen describe un experimento científico mucho mejor de lo que podría hacerlo un párrafo con muchas palabras:

¿Qué parte de la planta absorbe la MAYOR PARTE del agua?



- (A) Parte A
- (B) Parte B
- (C) Parte C
- (D) Parte D

5011026

Fuente: IEA 2007, ítem de ejemplo.

La diagramación de las opciones múltiples debe ser coherente. Se debe identificar cada opción. Se recomiendan las siguientes opciones de diagramación:

- Una columna vertical de opciones etiquetadas de arriba a abajo:

A

B

C

D

- Una línea horizontal de opciones etiquetadas de izquierda a derecha:

A

B

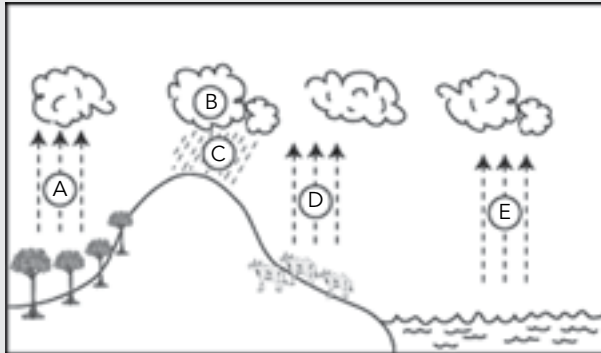
C

D

RECUADRO 3.16

Uso de imágenes simples

Esta imagen expresa un sistema complejo de un modo simple:



Fuente: Departamento de Educación de Filipinas 2004.

RECUADRO 3.17

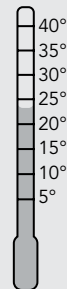
Gráficos con etiquetas claras

Las etiquetas claras y coherentes como las que aparecen en el siguiente gráfico ayudan a los estudiantes a entender rápidamente las preguntas complejas:

Esta tabla muestra las temperaturas a diferentes horas durante cuatro días.

¿Qué día y a qué hora se observó una temperatura en la tabla igual a la que se muestra en el termómetro?

Temperatura					
	6 a.m.	9 a.m.	Mediodía	3 p.m.	6 p.m.
Lunes	15°	17°	24°	21°	16°
Martes	20°	16°	15°	10°	9°
Miércoles	8°	14°	16°	19°	15°
Jueves	8°	11°	19°	26°	20°



Termómetro

- A. Lunes, mediodía
- B. Martes, 6 a.m.
- C. Miércoles, 3 p.m.
- D. Jueves, 3 p.m.

Fuente: Mullis et al. 2000.

RECUADRO 3.18**Mapas con etiquetas claras**

Se debe tener cuidado al etiquetar los elementos de un mapa y hacerlo de forma clara. En el siguiente mapa, todos los nombres de los continentes se muestran completamente en mayúscula, mientras que los océanos se muestran con mayúscula inicial y, luego, minúscula:



Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

- Dos columnas verticales de opciones, etiquetadas de arriba a abajo en la primera columna y, luego, de arriba a abajo en la segunda columna:

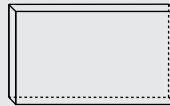
A	C
B	D

RECUADRO 3.19**Dejar espacios en el material de estímulo**

El espacio en este dibujo permite la legibilidad del material de estímulo:

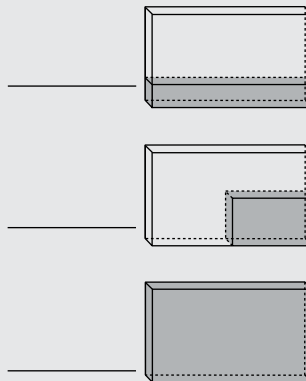


La figura anterior muestra una caja que contiene un material que podría ser un sólido, un líquido o un gas. A continuación, el material se pone dentro de una caja que es cuatro veces más grande.



Observe las imágenes a continuación. Muestran cómo se verán los diferentes tipos de material cuando se pongan en una caja más grande.

A. Identifique qué figura muestra un sólido, cuál muestra un líquido y cuál muestra un gas. (Escriba la palabra Sólido, Líquido o Gas en la línea al lado de cada figura a continuación. Utilice cada palabra solo una vez).



B. Explique sus respuestas.

5031372

Fuente: IEA 2003, ítem de ejemplo.

EL EQUIPO DE REDACCIÓN DE ÍTEMS

El encargado del desarrollo de la prueba dirige y gestiona el equipo que redacta los ítems y supervisa el programa completo, desde el desarrollo y la presentación ante el comité encargado de revisar los ítems, hasta la prueba previa y la selección de los formularios finales de la prueba. Debe tener un buen trato “personal” y habilidades organizativas.

Entre sus responsabilidades se incluyen las siguientes:

- Seleccionar al equipo de redactores de ítems.
- Garantizar que el equipo de redactores entienda el proyecto.
- Perfeccionar el proyecto.
- Establecer un conjunto de reglas o protocolos para presentar, clasificar y guardar los ítems.
- Garantizar que los redactores de los ítems estén al tanto de la cantidad de espacio en la página que pueden ocupar los ítems.
- Establecer y supervisar los procesos de presentación de los ítems ante el comité.
- Comparar el progreso del desarrollo de los ítems con el proyecto de prueba.
- Revisar los ítems con grupos de expertos o grupos de interés clave.
- Supervisar la calidad de los ítems.
- Hacer un seguimiento del desarrollo de los ítems con respecto a los calendarios establecidos.
- Anotar los detalles de todos los ítems desarrollados, entre ellos los antecedentes de la prueba previa y los cambios hechos durante el análisis.
- Garantizar que el proyecto se refleje en la prueba final.

La redacción de los ítems requiere de atención al detalle, creatividad, rigor intelectual, profundidad de conocimiento del contenido y un buen entendimiento del desarrollo de los estudiantes en un área de aprendizaje. Es ideal que los redactores de ítems demuestren las siguientes características:

- Deben mostrar iniciativa y predisposición a realizar una investigación extensa en busca de materiales de estímulo interesantes, y deben tener la capacidad de desarrollar materiales de estímulo de alta calidad.

- Deben tener la capacidad de aceptar comentarios sobre su trabajo y de realizar comentarios sobre el trabajo de otros redactores de ítems con el mismo grado de imparcialidad crítica.
- Deben demostrar el deseo de lograr la excelencia y la predisposición a estar atentos a los detalles al desarrollar y perfeccionar los ítems.

Vale la pena considerar una prueba de selección inicial al seleccionar a los redactores de ítems. La prueba puede consistir en darles a los potenciales redactores 30 minutos para generar ítems que se basen en un conjunto de materiales de estímulo. Esta prueba puede estar seguida de una entrevista durante la cual se les pida a los candidatos que expliquen razonadamente sus respuestas a la prueba de selección. El comité entrevistador puede verificar la voluntad de los potenciales redactores de ítems de aceptar críticas sobre su trabajo.

Es ideal que algunos redactores de ítems con experiencia participen en la capacitación de los nuevos redactores de ítems. Estos redactores con experiencia pueden provenir de otro país y haber sido contratados como asesores si no hay nadie disponible con la experiencia adecuada dentro del país. Los redactores de ítems que son asesores pueden realizar sesiones de capacitación, revisar los ítems mientras se desarrollan o ambas cosas. A los redactores de ítems que trabajen a tiempo completo luego de la capacitación les puede llevar varios meses alcanzar el punto en el que comiencen a producir ítems de una calidad razonable.

En la capacitación se deben abordar las siguientes preguntas:

- ¿Cuál es el propósito general de la prueba?
- ¿Qué contextos de estímulo son apropiados para los ítems?
- ¿Qué aspectos del currículo cubrirá la prueba?
- ¿Qué proporción de ítems cubrirán los diferentes aspectos del currículo?
- ¿Qué idioma (o idiomas) se usarán?
- ¿Cuál es el nivel apropiado de simplicidad en el vocabulario y la gramática utilizados?
- ¿Qué formatos de ítem se usarán y en qué proporción?
- ¿Cuáles son las especificaciones de la publicación (número de páginas del cuadernillo de la prueba, tamaño de la página, número de ítems por página)?

- ¿Cuántos ítems se proponen para la versión final de la prueba?
- ¿Cuántos ítems se deben desarrollar?
- ¿Cómo se revisarán los borradores de los ítems de la prueba?
- ¿Cuál es el margen de tiempo para el desarrollo, la prueba previa y la selección de los formularios finales?
- ¿Se deben tener en consideración asuntos culturalmente sensibles o restricciones al redactar el borrador del material de estímulo y los ítems?

Los redactores de los ítems deben tener un entendimiento común de las respuestas a estas preguntas. También deben supervisar regularmente su propio trabajo y el de los demás. Todos los redactores de ítems deben tener copias de las especificaciones técnicas de la prueba terminadas, así como un entendimiento común de sus contenidos.

La especificación en detalle del estilo de la presentación de ítems desde el principio ahorra mucho tiempo a largo plazo. El encargado del desarrollo de la prueba debe establecer una hoja de estilo que especifique exactamente cómo se presentarán los ítems y las guías de corrección. La hoja de estilo debe cubrir todos los aspectos de diagramación, entre ellos la selección de fuentes y su tamaño, el uso de sangrías, la disposición de los encabezados y los tipos de detalles que se deben incluir, como se muestra en el ejemplo en el recuadro 3.20.

Esta guía muestra que los redactores de ítems deben darle un título a su unidad en Arial negrita de 16 puntos con mayúscula. El resto del texto para el ítem está en Times New Roman. La mayor parte está escrita en 12 puntos. La pregunta se debe etiquetar, en negrita, como “Pregunta 1”. El nombre de la unidad debe aparecer de la siguiente forma. El ítem y el espacio para la respuesta del alumno se encuentran debajo de la pregunta. El redactor del ítem inserta y completa una tabla para mostrar el tipo de texto, el formato del ítem y los procesos que utilizan los alumnos para responder el ítem. La guía de puntuación se etiqueta de la siguiente forma. Los criterios para el puntaje se muestran en cursiva, con ejemplos de respuestas de alumnos presentados como viñetas con sangría en 10 puntos. Al seguir la guía, los redactores del ítem pueden ayudar a garantizar que los ítems de la prueba previa y los finales estén preparados de un modo coherente, general y eficiente.

RECUADRO 3.20**Ejemplo de hoja de estilo para los redactores de ítems****TRADICIONES DE PASCUA (título, 16 puntos, Arial negrita)**

Pregunta 1: Tradiciones de Pascua (subtítulo, 12 puntos, Times New Roman negrita)

¿Qué regalos se dieron el domingo de Pascua? (pregunta, 12 puntos, Times New Roman negrita)

<insertar media línea> (instrucción para la publicación en cursiva y entre paréntesis)

Tipo de texto	Formato del ítem	Proceso
Información	De respuesta cerrada	Recuperación

(tabla con 3 columnas y 2 filas: títulos de la columna, 12 puntos, Times New Roman negrita; cuerpo de la tabla 12 puntos, Times New Roman sin negrita)

Guía de corrección (subtítulo, 12 puntos, Times New Roman negrita)

1 punto: *se refiere a los huevos* (12 puntos, Times New Roman cursiva)

- Se regalaron huevos. (viñeta, 10 puntos, Times New Roman)
- Decoraron los huevos.

0 puntos: *se refiere a panqueques, otros o respuesta vaga*

- Panqueques.
- Se regalaron cosas.

Los redactores de ítems deben recibir comentarios periódicos y claros, y supervisión constructiva en cuanto a sus propios ítems y el modo en que coinciden con las especificaciones técnicas, para que puedan aprender de sus errores, desarrollar sus habilidades y perfeccionar sus ítems. Los redactores de ítems deben reunirse con regularidad y con frecuencia en comités encargados de revisar los ítems a fin de criticar su trabajo. El encargado de la prueba debe estar preparado para reemplazar a los redactores que no puedan desarrollar ítems de alta calidad luego de un período de capacitación razonable.

COMITÉS ENCARGADOS DE REVISAR LOS ÍTEMS

Un comité encargado de revisar los ítems consta de un pequeño grupo (de entre tres y seis) redactores de ítems que revisan conjuntamente el material que uno o más de ellos han desarrollado. El objetivo del comité es aceptar, modificar o rechazar el material. El enfoque de equipo, que es parte del proceso de control de calidad, ayuda a obtener múltiples perspectivas sobre ítems individuales. A menos que los redactores de ítems tengan mucha experiencia, los ítems generalmente se someten a una revisión significativa luego de la revisión hecha por el comité.

Los miembros del comité deben preparar sus críticas antes de que se reúna el comité encargado de revisar los ítems. Deben tener abundante tiempo para examinar los ítems y hacer un borrador de las mejoras sugeridas.

El comité debe someter a crítica cuidadosa el contexto del material de estímulo, el contenido, la redacción, el lenguaje, el esquema y las ilustraciones para garantizar que cada aspecto del estímulo sea relevante para el proyecto, su inclusión en la prueba valga la pena y sea claro y conciso. A continuación, debe examinar exhaustivamente cada ítem para garantizar que la redacción no sea ambigua, que el formato sea adecuado y que el ítem aborde con claridad las habilidades y áreas de contenido especificadas en el proyecto. El conjunto de ítems también se examina para considerar el grado en que el equilibrio general de los ítems refleja las especificaciones técnicas. Los miembros del comité deben explorar cada posibilidad de mejorar el estímulo y los ítems y, cuando sea necesario, hacer sugerencias de nuevos ítems.

Durante el proceso del comité encargado de revisar los ítems, los redactores de ítems deben explicar su trabajo y estar preparados para aceptar críticas constructivas. El líder del comité debe garantizar que haya un consenso amplio sobre los cambios que se van a realizar en los ítems individuales. Los redactores de los ítems deben documentar los cambios sugeridos y, posteriormente, revisar los ítems.

Puede ser necesario un experto en idiomas en los comités donde los redactores de ítems estén redactando pruebas en una lengua que no es la materna. El experto en idiomas debe tener una buena comprensión de las aptitudes lingüísticas de la población objetivo de la prueba.

En el comité se puede incluir a un experto en el área, en especial si el área del contenido es compleja. La participación de un experto en algunos comités puede ser útil para aclarar los problemas de contenido, pero no es necesario que esta participación sea continua. Es poco probable que el experto tome parte en los puntos más delicados de la redacción del ítem.

Es aconsejable no incluir a responsables políticos ni grupos de interés clave en los comités encargados de revisar los ítems. No les conciernen los detalles más técnicos de las deliberaciones del comité.

Los miembros del comité deben tener en cuenta cada aspecto de un ítem:

- ¿Se evalúa el contenido correcto?
- ¿El formato del ítem es adecuado para los estudiantes objetivo?
- El ítem, ¿es importante o es trivial?
- ¿La redacción del ítem es clara e inequívoca?
- ¿Hay errores ortográficos o se omiten palabras?
- Si el ítem es de opción múltiple, ¿las opciones son similares y significativas?
- Si el ítem es de opción múltiple, ¿la opción correcta se deduce con claridad y de forma inequívoca a partir de la información dada (el enunciado, el material de estímulo o ambos)?
- ¿El esquema del ítem es atractivo y ordenado?
- La mayoría de los ítems, ¿presentan alguna dificultad que le permitirá a entre el 40 % y el 80 % de los estudiantes evaluados dar la respuesta correcta?
- Si el ítem requiere de un puntaje de crédito parcial, ¿es probable que al menos el 10 % de quienes responden consiga tal puntaje?
- ¿El ítem parece ser imparcial y justo para la mayoría de los subgrupos de la población objetivo?

El comité también debe considerar las siguientes formas de mejorar el ítem:

- Acortarlo.
- Agregar más información.
- Cambiar una expresión o la redacción.
- Agregar un diagrama o una ilustración.
- Reformularlo en un formato de ítem diferente.

Los redactores de ítems deben recibir comentarios periódicos y frecuentes desde el momento en que comienzan a desarrollarlos. Los comités encargados de revisar los ítems se deben reunir una vez por semana, si es posible.

La revisión de los materiales puede ser una tarea compleja. El comité necesita un líder que garantice que las recomendaciones sean unánimes y que se alcance un consenso sobre los cambios que se van a realizar. El redactor de ítems no es la persona correcta para decidir qué cambios se adoptan o se dejan de lado. Las recomendaciones del líder del comité deben ser lo suficientemente precisas como para garantizar que el redactor del ítem tenga en claro qué cambios hacer.

Rehacer los ítems de otras personas puede suscitar fuertes debates en algunos casos. El líder del comité debe centrar la discusión en la mejora del ítem y garantizar que el comité trabaje de forma constructiva hacia este objetivo.

Se debe revisar el trabajo de todos los miembros del comité, entre ellos los redactores de ítems con experiencia. Es normal, en especial para los nuevos redactores de ítems, esperar sugerencias de revisiones extensas. La crítica de los ítems no se debe ver como una crítica hacia el individuo. Se debe reemplazar a los redactores que no acepten la discusión a fondo y la reformulación de sus ítems.

OTROS REVISORES

Los grupos de expertos o los grupos de interés clave deben tener la oportunidad de revisar el conjunto de ítems varias veces durante su desarrollo. Este procedimiento puede ayudar a garantizar que los ítems de la prueba sean de buena calidad y coherentes con las especificaciones técnicas. Por lo general, el encargado del desarrollo de la prueba presenta todos los ítems, o una selección, ante un grupo de referencia seleccionado para este propósito.

La primera revisión con el grupo de referencia debe llevarse a cabo en un momento razonable a comienzos del proceso de desarrollo de los ítems, a fin de garantizar que los redactores estén trabajando en la dirección correcta. El grupo de referencia puede sugerir el ajuste de algunos aspectos del proyecto, en especial si los redactores tienen

problemas para cumplir algunas de las especificaciones. Además, los redactores de ítems tal vez necesiten indicaciones más específicas sobre el material aceptable e inaceptable.

Normalmente, se realiza una revisión luego de finalizado el borrador de todos los ítems, para garantizar que los grupos de interés aprueben los ítems antes de que se lleve a cabo la prueba previa. Una revisión final permite que los grupos de interés clave aprueben la selección de ítems para el formulario de prueba final.

SEGUIMIENTO DE LOS ÍTEMS

Es crítico hacer un seguimiento de los ítems. Cada ítem debe tener una etiqueta única para que se le pueda hacer un seguimiento en cada etapa, desde la prueba previa hasta el análisis. Se deben desarrollar aproximadamente entre más del doble y el triple de ítems de los que se requieren para los formularios de examen finales. Por lo general, es necesario producir varios cuadernillos de cada nivel escolar para la prueba previa. En los diferentes cuadernillos deben aparecer algunos de los mismos ítems. Esto permite que todos los ítems previamente examinados se relacionen en la misma escala y se comparen. La etiqueta de cada ítem debe ser independiente del orden del ítem en el cuadernillo, de modo que los ítems duplicados en diferentes cuadernillos y los ítems únicos se puedan identificar con claridad.

El seguimiento de los ítems durante el análisis puede ser complicado. Si el analista omite darle una etiqueta única a un ítem, puede ser muy difícil hacerle un seguimiento. Los programas informáticos pueden agravar el problema. Con frecuencia, los programas numeran los ítems de forma automática. Si los ítems se toman del análisis de la prueba previa, el programa volverá a numerarlos, por lo que quizás el número de ítem en el análisis ya no coincida con su número en el análisis inicial o en el cuadernillo de examen. Darle a cada ítem la misma etiqueta única en el cuadernillo de prueba y en cada uno de los análisis ayudará a evitar este problema.

Las etiquetas de los ítems deben ser lo más significativas posible. El encargado del desarrollo de la prueba debe coordinarse con el analista de datos para establecer cuántos caracteres se pueden utilizar para

una etiqueta. Los paquetes de programas para análisis tienen diferentes límites. Una agencia de evaluación utiliza la siguiente convención de etiquetado para una prueba de comprensión lectora, escritura y matemáticas administrada durante varios años:

- El primer carácter es C, M o E, por comprensión lectora, matemáticas o escritura.
- Los caracteres 2 y 3 indican el año en que se administró el ítem (por ejemplo, 07 para 2007).
- Los caracteres 4, 5 y 6 representan el ítem (comenzando con 001).

Por ejemplo, M06003 indica que el ítem es un ítem de una prueba de matemáticas administrado en 2006 y que es el tercer ítem en el conjunto de ítems.

Los ítems se deben etiquetar durante el desarrollo. En los cuadernillos de prueba previa y en los formularios finales se debe imprimir la misma etiqueta. Las etiquetas se pueden imprimir en una fuente pequeña en escala de grises sobre el margen opuesto a los números de ítem en los cuadernillos de prueba, como se muestra a continuación:

M06003 5 Complete esta suma.
6 + 7 = _____

Mantener un historial completo de cada ítem que se desarrolle es esencial. El analista de datos necesita un registro de las claves para los ítems de opción múltiple y la ubicación de los ítems en los cuadernillos de prueba previa. Con frecuencia, los informes requieren información sobre el formato de los ítems y los procesos evaluados por cada uno de ellos. El encargado del desarrollo de la prueba debe preparar y conservar una hoja de cálculo como registro de cada ítem, sus clasificaciones y su estado, y todo cambio que se pueda haber hecho al ítem.

El siguiente es un ejemplo de algunos de los encabezados de columna en un registro de hoja de cálculo de todos los ítems de comprensión lectora para una prueba:

- **nombre de la unidad** nombre dado a la unidad (por ejemplo, “Autos de carrera”)
- **etiqueta del ítem** etiqueta de 6 dígitos (por ejemplo, C06003)

- **contenido del ítem** redacción de la pregunta de la prueba
- **estado actual** descripción que indica si el ítem está disponible para su uso (por ejemplo, publicado como ítem de práctica, rechazado por el cliente, permiso de derechos de autor rechazado)
- **clave** respuesta correcta para una pregunta de opción múltiple
- **puntaje máx.** puntaje máximo para el ítem
- **tipo de texto** género del texto (por ejemplo, narrativo o informativo)
- **formato del ítem** formato de la pregunta de la prueba (por ejemplo, de opción múltiple)
- **proceso** proceso cognitivo (por ejemplo, recuperación)
- **notas del análisis** cambios realizados al ítem luego de la prueba previa

Los ítems se deben guardar en un lugar seguro. Todo material relevante relativo al desarrollo de una unidad o un ítem se debe guardar con ese ítem. Incluso se debe conservar el material que no se utilice en la prueba previa, dado que se puede utilizar más tarde en el mismo nivel escolar o en otro. La fuente de los documentos o las ilustraciones se debe registrar y guardar con la unidad o el ítem, de modo que se pueda solicitar el permiso de derechos de autor si es necesario. Se deben conservar copias de los documentos originales para que se pueda identificar toda modificación posterior.

La mayoría de los ítems se pueden guardar en formato electrónico. Como precaución, consérvense archivos de seguridad de los ítems en una computadora o disco aparte. El etiquetado correcto y la clasificación total y precisa ayudan a garantizar que los ítems se almacenen en las carpetas de archivos (de computadora) apropiadas y que otros los puedan recuperar con facilidad. Los ítems suelen estar bajo revisión constantemente, e incluso sufrir cambios en las guías de corrección y las ilustraciones, así como mejoras mínimas en la redacción. La última versión del ítem debe destacarse en la información de la carpeta de archivos.

NOTAS

1. Para mayor información sobre la estructura de los ítems de prueba, consulte Chatterji (2003), Haladyna (1999), Kubiszyn y Borich (2000) y Linn y Miller (2004).
2. La Evaluación Nacional del Progreso Educativo de los EE. UU. incluye un componente de escritura (Consejo Directivo de Evaluaciones Nacionales, sin fecha).

ENSAYO DE LOS ÍTEMS

Para elaborar pruebas de evaluación nacionales se utiliza, en gran medida, la tecnología desarrollada durante el diseño de las pruebas destinadas a evaluar e informar sobre el rendimiento académico de estudiantes individuales. Dado que tales pruebas se utilizan para discriminar entre los desempeños de los estudiantes, todos ellos realizan esencialmente la misma prueba. El propósito de una evaluación nacional es muy diferente: no pretende discriminar entre los estudiantes, sino describir el grado en que los estudiantes de un sistema educativo en su conjunto (o una parte claramente definida del mismo) han adquirido los conocimientos y habilidades establecidos en un plan de estudios. Para ello, la prueba deberá proporcionar una cobertura curricular adecuada, lo cual puede requerir una muestra del plan de estudios mucho mayor que la necesaria en pruebas diseñadas para evaluar a estudiantes individuales. La necesidad de una cobertura curricular amplia se intensifica cuando la evaluación pretende identificar las áreas del currículo en las que los estudiantes presentan puntos fuertes y débiles.

Para abordar estas cuestiones, numerosas evaluaciones nacionales e internacionales utilizan un número de ítems mucho mayor que en las pruebas diseñadas para evaluar a estudiantes individuales. Sin embargo,

para evitar imponer una carga demasiado pesada sobre los estudiantes, cada uno de ellos responde únicamente a una fracción del número total de ítems de la evaluación. Por lo tanto, se deben proporcionar varios conjuntos alternativos de ítems en un cuadernillo de diseño rotativo. El número exacto varía de una evaluación nacional a otra.

Este enfoque, aunque deseable en muchos sentidos, conlleva una serie de complicaciones para la administración de una evaluación nacional. En primer lugar, el diseño de la prueba es complejo, ya que se debe garantizar la superposición entre preguntas y la coincidencia entre submuestras. En segundo lugar, la administración es más compleja, dado que es necesario garantizar que los cuadernillos se entreguen a los estudiantes correctos y que las instrucciones proporcionadas a los estudiantes sean adecuadas para todos los cuadernillos. Por último, la combinación de datos de varios conjuntos de ítems requiere utilizar procedimientos estadísticos relativamente complejos. Por estas razones, muchos países en desarrollo no han utilizado cuadernillos de prueba múltiples en sus evaluaciones nacionales.

La mayoría de los comentarios en las páginas siguientes, y también en el capítulo 5, son válidos tanto para situaciones en las que un equipo de evaluación nacional utiliza cuadernillos de prueba múltiples como para aquellas en las que se opta por un único cuadernillo para medir los logros de aprendizaje en un área temática. Ambos enfoques obligan a asegurarse de que las pruebas previas se realicen de forma cuidadosa.

Las pruebas previas o pruebas piloto son un elemento esencial del desarrollo de pruebas. Las pruebas previas se administran a alumnos que tengan las mismas características que los que van a participar en la prueba final. Se deben incluir escuelas de diferentes tamaños, en diferentes áreas y con estudiantes de diferentes niveles socioeconómicos. Idealmente, la prueba previa se debe llevar a cabo con un año de antelación, en la misma época del año que la prueba final programada. Por ejemplo, si la prueba previa se realiza con estudiantes de quinto grado en octubre de 2005, la prueba final se debe realizar con estudiantes de quinto grado en octubre de 2006. En la práctica, esto puede no ser posible, y puede que haya que realizar la prueba previa con estudiantes que tienen un par de meses más o menos de experiencia escolar que el grupo objetivo. Por ejemplo, se puede realizar la prueba

previa con estudiantes de sexto grado al principio del año escolar para obtener datos para una prueba que se va a administrar a estudiantes de quinto grado al final del año escolar.

Los estudiantes que participen en la prueba previa no deben realizar la prueba final. Si la prueba final se va a administrar a un grupo de muestra aleatorio seleccionado estadísticamente, el grupo de muestra final se debe seleccionar antes de elegir las escuelas para la prueba previa.

Una buena práctica es realizar una prueba previa con un número de ítems dos o tres veces mayor que el necesario para la prueba final. Los formularios de la prueba previa deben tener una duración similar a la de la prueba final. Es de gran ayuda etiquetar los formularios con el grado en orden alfabético; por ejemplo, si se dispone de cinco formularios, los del tercer grado se identificarían como 3A, 3B, 3C, 3D y 3E, mientras que los del octavo grado se identificarían como 8A, 8B, 8C, 8D y 8E.

Se requieren varios formularios de prueba previa para cada grado. Lo ideal sería distribuir los formularios de grado al azar en cada clase. Si, por ejemplo, se utilizan tres formularios para el quinto grado (5A, 5B y 5C), cada escuela debe recibir una mezcla de los tres. Si no es posible seguir este procedimiento, es importante asegurarse de que cada modelo de formulario se distribuya de forma transversal en toda la muestra de la prueba previa. Por ejemplo, el formulario 5A no se debe entregar únicamente a alumnos urbanos, el 5B a alumnos de zonas rurales del norte y el 5C a alumnos de zonas rurales del sur. Debe haber la máxima equivalencia posible entre los distintos formularios en lo que respecta a las especificaciones técnicas de la prueba.

Es esencial que los formularios estén vinculados entre sí para poder agrupar los ítems y compararlos. La vinculación significa que algunos ítems se repiten en formularios distintos. Es inevitable que algunos formularios de la prueba previa sean más difíciles que otros. La vinculación de los formularios permite determinar la dificultad global de los ítems, sin importar el formulario en el que aparecen. La vinculación horizontal solo es necesaria en pruebas para un único grado, mientras que para pruebas en varios grados se requiere vinculación vertical.

Cada uno de los ítems de la prueba previa debe ser respondido por un mínimo de 200 estudiantes de cada grado. Si, por ejemplo, una prueba previa dispone de tres formularios para el quinto grado, al menos 600 estudiantes deben participar en la prueba. Se necesitan al menos 150 respuestas de los 200 estudiantes para cada ítem de la prueba previa. Es inevitable que se pierdan algunos datos durante una prueba previa. Es posible que, por motivos justificados, una escuela se retire del programa de pruebas previas en el último momento o que administre la prueba a menos estudiantes de los previstos. Además, es improbable que todos los estudiantes respondan a todos los ítems de la prueba previa.

Debido a que las pruebas previas se deben llevar a cabo en las mismas condiciones que la prueba final, el tiempo del que los estudiantes disponen para realizar la prueba debe ser el mismo que en la prueba final. Es posible que no se conozca el número de ítems que los estudiantes son capaces de completar durante ese tiempo. Si es así, elabore una prueba previa de muestra y ensáyela en algunas clases antes de preparar la prueba previa, para asegurarse de que el número de ítems de cada formulario de la prueba previa sea realista. Una evaluación nacional no es una prueba de velocidad. La mayoría de los estudiantes deben disponer de tiempo suficiente para responder a la mayoría de los ítems.

Intente completar todas las pruebas previas a lo largo de un periodo de dos o tres semanas.

Las pruebas previas permiten evaluar la idoneidad y la calidad de los ítems. También permiten perfeccionar numerosos aspectos de la administración de las pruebas. Se debe solicitar a los administradores de las pruebas previas que proporcionen la información solicitada a continuación:

- ¿Han tenido los estudiantes suficientes preguntas de práctica, y han sido las instrucciones y explicaciones lo suficientemente claras?
- ¿La duración de la prueba ha sido adecuada o excesiva, y cuántos estudiantes aproximadamente han finalizado la prueba diez minutos antes de agotarse el tiempo disponible?
- ¿Parecían los estudiantes estar inmersos en la prueba?
- ¿Contaban los estudiantes con suficientes recursos, tales como lapiceros y gomas de borrar?

- ¿Son adecuadas las instalaciones escolares para realizar una prueba?
- ¿Han comprendido los profesores y los estudiantes el propósito de la prueba?

DISEÑO DEL FORMULARIO DE LA PRUEBA PREVIA

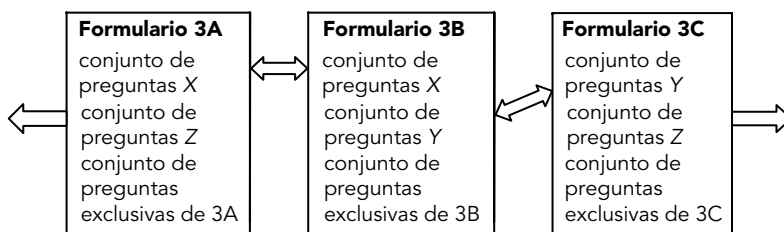
El análisis de los datos de la prueba previa sirve de base para seleccionar ítems para la prueba final. En muchas evaluaciones nacionales se elaboran modelos distintos de cuadernillos de prueba vinculados para cada grado. Este enfoque proporciona una cobertura curricular mayor que una única prueba y, al mismo tiempo, contribuye a garantizar que los estudiantes no sean sometidos a pruebas desmesuradamente largas. Las pruebas se deben diseñar y vincular correctamente para asegurarse de que los datos se puedan combinar en una única escala. Los analistas de datos, estadísticos o informáticos profesionales deben participar en el diseño de la prueba previa para ayudar a garantizar que se cumplan los requisitos relativos a los datos.¹

Los formularios vinculados comparten algunos ítems. Por lo general, se necesitan entre ocho y diez ítems comunes. Hay varias maneras de vincular los formularios. Cuando se trata de un *conjunto único de ítems comunes de enlace*, se repiten los mismos ocho o diez ítems de enlace en cada formulario. Se debe tener en cuenta que si el resultado del análisis de los ítems de enlace no es bueno (características estadísticas mediocres), la vinculación entre los formularios será débil y, por consiguiente, la calidad del análisis global se verá perjudicada.

Una segunda manera es la *vinculación circular*. Se utilizan conjuntos distintos de ítems entre parejas de formularios. Por ejemplo, se puede vincular el formulario 3A al formulario 3B con el conjunto de ítems X, el formulario 3B al formulario 3C con el conjunto de ítems Y y el formulario 3C de nuevo al formulario 3A con el conjunto de ítems Z. De esta forma, cada formulario contiene ítems exclusivos que no aparecen en ninguno de los otros formularios (figura 4.1).

Una tercera forma es la *vinculación lineal*, que sigue el modelo de la vinculación circular, pero excluyendo los vínculos del conjunto de preguntas Z. Por lo tanto, el formulario 3A estaría vinculado al 3B y el

FIGURA 4.1

Ejemplo de vinculación circular de ítems

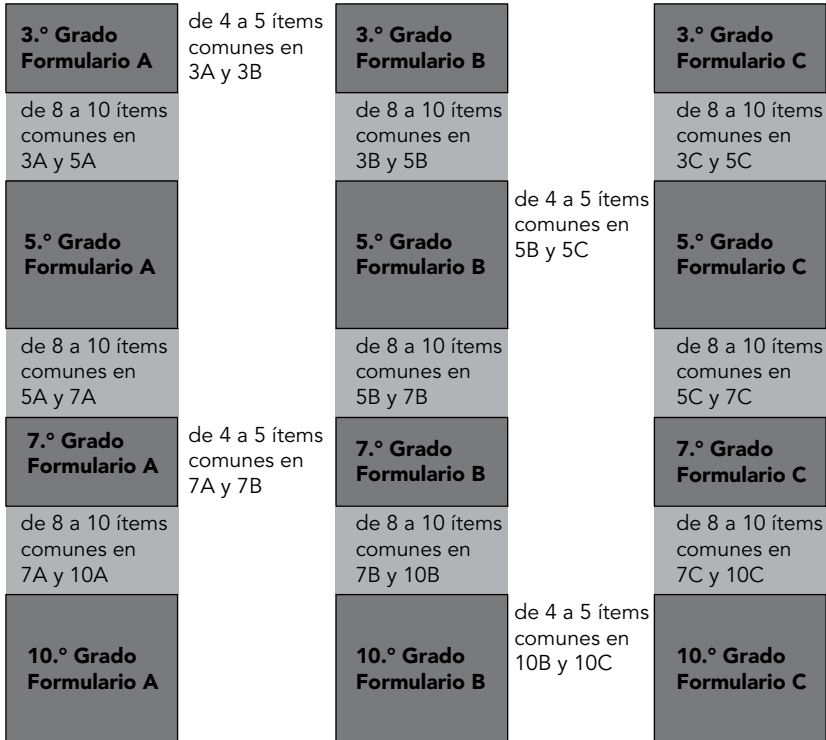
formulario 3B al 3C, pero no se establecería una vinculación entre los formularios 3A y 3C.

Si no está claro el resultado que se obtendrá con los ítems, se recomienda utilizar la vinculación circular y más ítems de enlace de los necesarios. En la vinculación circular se conservan los vínculos entre los formularios, incluso si un conjunto de ítems de enlace falla.

La vinculación puede ser tanto horizontal (dentro de un grado) como vertical (en varios grados; por ejemplo, entre los grados 3.º y 5.º). Si los formularios de la prueba final no se van a vincular verticalmente, entonces la prueba previa debe contar con una vinculación horizontal profunda. También se puede incluir una vinculación vertical mínima para permitir la comparación entre datos de la prueba previa de grados distintos. Técnicamente, solo son necesarios entre ocho y diez ítems comunes a los distintos grados. Es más difícil determinar cómo debe ser una pregunta con una buena vinculación vertical que con una buena vinculación horizontal. Debido a que se trata de una prueba previa y se desconoce la calidad de los ítems de enlace, se recomienda incluir al menos dieciséis ítems de enlace vertical distribuidos en cada uno de los formularios. Si los formularios definitivos se deben vincular verticalmente, es importante que el ensayo cuente con más ítems de enlace vertical de los necesarios en la prueba final, para poder seleccionar los mejores ítems de enlace para el formulario definitivo. En la figura 4.2 se muestra un modelo de prueba previa alternativo con formularios finales vinculados verticalmente. El modelo se basa en tres formularios—A, B y C—distribuidos aleatoriamente dentro de cada clase.

FIGURA 4.2

Modelo para ítems de vinculación vertical



En este ejemplo bastante elaborado, cada uno de los formularios 3A, 5A, 7A y 10A está vinculado verticalmente de forma lineal mediante entre 8 y 10 ítems. Los ítems están vinculados de manera similar en los formularios B y C. En total, hay entre 8 y 10 ítems de enlace horizontal entre los formularios A y B en el 3.º grado y el 7.º grado, y entre 8 y 10 ítems de enlace horizontal entre los formularios B y C en el 5.º grado y el 10.º grado. Este número de vínculos horizontales es aceptable. Si los formularios no se van a distribuir aleatoriamente en cada clase, o si los redactores de los ítems no están seguros de la calidad de los ítems de enlace horizontal, se deberían incluir más vínculos horizontales en cada grado.

Es preferible situar los ítems de enlace en la parte inicial o central de los formularios de las pruebas, en lugar de la parte final, para evitar

la posibilidad de que los estudiantes no respondan a estos ítems. Los ítems de enlace se deben colocar en un orden similar y posiciones similares en cada uno de los cuadernillos para que no sea posible atribuir diferencias en el rendimiento de los alumnos a diferencias en la posición o el orden de los ítems. Es inevitable que existan pequeñas diferencias en la ubicación de los ítems de enlace. Se deben evitar grandes diferencias.

La dificultad de los ítems de enlace se debe situar en el promedio. En general, los estudiantes de capacidad promedio de la población objetivo deben tener una probabilidad de entre el 40 y el 60 % de responder correctamente a estos ítems. Debido a que se trata de una prueba previa, no se conocerá la dificultad de los ítems para la población de la prueba previa; por ello, los redactores de los ítems tendrán que estimar su dificultad lo mejor que puedan. Si los redactores de los ítems no están seguros de la precisión de sus estimaciones, se recomienda aumentar el número de ítems de enlace.

Si los ítems están dispuestos en unidades, es mejor vincular los formularios mediante ítems extraídos de dos o más unidades, en caso de que los ítems asociados a una unidad no den buen resultado. No es necesario utilizar todos los ítems de una unidad para la vinculación; algunos ítems pueden ser comunes y otros exclusivos, como se muestra en la tabla 4.1. Ambas unidades—«Encontrar una mascota» y «Monte Avarapa»—aparecen en los formularios 3A y 3B del 3.º grado. Tres de los ítems son comunes a todas las unidades y todos los formularios. La prueba previa tiene ítems adicionales exclusivos.

Si la prueba previa incluye una combinación de formatos de ítems, los ítems de enlace también deben reflejar esta combinación.

Los ítems deben estar acompañados de etiquetas únicas impresas en escala de grises en todos los formularios en los que aparecen.

TABLA 4.1
Ítems de enlace en dos unidades de lectura

Unidad de lectura	Ítems comunes a 3A y 3B	Ítems exclusivos de 3A	Ítems exclusivos de 3B
«Encontrar una mascota»	3, 4, 6	2, 7	1, 5
«Monte Avarapa»	1, 2, 5	3	4

TABLA 4.2
Parte de una hoja de cálculo para controlar los ítems en los distintos formularios

Nombre de la unidad	Etiqueta de la pregunta	Formulario 5A	Formulario 5B	Formulario 5C	Formulario 5D
«Perros»	R070101	1	1	4	4
«Perros»	R070102	3	3	5	5
«Perros»	R070103		2		
«Perros»	R070104	2			
«Ellie»	R070201			1	
«Ellie»	R070202			2	1
«Ellie»	R070203			3	2
«Ellie»	R070204				3
«Bang»	R070301	4			6
«Bang»	R070302	5			7

Los ítems con la misma etiqueta deben ser completamente idénticos, salvo por el orden de aparición en el formulario de prueba. Los ítems con una formulación ligeramente distinta deben tener etiquetas distintas.

El redactor de los ítems debe crear una hoja de cálculo con una lista de todos los ítems; los encabezados deben indicar en qué formularios y en qué orden aparecen los ítems. La tabla 4.2 muestra un ejemplo de hoja de cálculo que cubre tres unidades («Perros», «Ellie» y «Bang») de una prueba de lectura del 5.º grado. A la izquierda se muestran los ítems para tres unidades. Los números de las columnas indican el orden de aparición de cada uno de los ítems en los formularios de la prueba previa. Los primeros dos ítems de «Perros» son comunes a los cuatro formularios.

IMPRESIÓN Y REVISIÓN DE LA PRUEBA PREVIA

El aspecto de los ítems en la prueba previa debe ser idéntico a su aspecto en el formulario definitivo. Igualmente, el material de estímulo, los gráficos y las ilustraciones se deben presentar de la misma

forma que aparecerán en la prueba final. Lo ideal es que el orden de aparición de los ítems de enlace también sea idéntico, pero, en la práctica, puede variar ligeramente.

El material de estímulo para los ítems de lectura debe aparecer (a) en la misma página que los ítems o (b) en la página izquierda, con los ítems en la página derecha para permitir a todos los alumnos desplazarse fácilmente entre los ítems y el texto.

La portada del cuadernillo de la prueba previa no necesita tener toda la información requerida para el formulario definitivo. Se debe solicitar información del estudiante como escuela, grado, sexo, lengua materna y edad. Debido a que, por lo general, no se realizan informes de las pruebas previas, normalmente no es necesario que los estudiantes incluyan su nombre en la prueba previa. Es posible que en la prueba previa no se solicite alguna información de referencia que sí será necesaria en el formulario definitivo. El diseño de los ítems debe ser coherente en todos los formularios.

La siguiente lista de control puede servir de ayuda para preparar o corregir la versión impresa de los materiales de la prueba previa:

- Encabezados (grandes y claros)
- Márgenes: superior, inferior, izquierda y derecha (uniformes)
- Números de página (uniformes)
- Números de ítem (grandes y claros)
- Etiquetas de ítems (presentes)
- Líneas para que los estudiantes escriban sus respuestas (claras y de longitud adecuada)
- Texto del ítem (tamaño de fuente de 12 o 14 puntos)
- Número de palabras por línea (de 10 a 12)
- Material de estímulo (claro, preferentemente en una fuente distinta a la de los ítems)
- Material de estímulo e ítems asociados (en la misma página o páginas opuestas)
- Encabezados y pies de página (uniformes y útiles)
- Revisión ortográfica (realizada).

Algunas pruebas incluyen opciones de calificación en escala de grises. Por ejemplo, se puede asignar un valor de 0 o 1 a una pregunta

para calificarla como incorrecta o correcta, respectivamente. Normalmente, la opción de calificación para ítems no respondidos es 9, como se indica en el capítulo 3. Los correctores simplemente deben trazar un círculo en torno a la puntuación adecuada. Las puntuaciones incluidas sirven para recordar a los correctores las opciones de calificación.

Todos los cuadernillos de pruebas y manuales de administración se deben revisar a fondo. Los revisores deben responder a los ítems de la prueba como si estuvieran realizándola. También deben asegurarse de que los materiales cumplan los siguientes criterios:

- Instrucciones iniciales e ítems de práctica claros e inequívocos.
- Ítems claros e inequívocos.
- Material de estímulo claro y fácil de leer.
- Opciones de respuesta múltiple con una respuesta correcta y las demás claramente incorrectas.
- Todas las opciones de respuesta múltiple deben tener sentido.
- Espacio suficiente para que los estudiantes anoten las respuestas, si es necesario.
- Material de estímulo para la lectura en la misma página que los ítems o en la página de la izquierda, con los ítems en la página opuesta, a la derecha.
- Los ítems de una unidad son independientes; es decir, la respuesta a un ítem no se encuentra en el enunciado o en las opciones de otro ítem.
- Los ítems de enlace son idénticos.
- No hay errores ortográficos ni gramaticales.
- El diseño de los distintos formularios de prueba es uniforme.

La revisión es esencial. Se pueden producir, y se producen, graves errores en prácticamente todas las fases del proceso de las pruebas previas. Una prueba previa con errores tipográficos e incoherencias representa una tremenda pérdida de tiempo, esfuerzo y dinero. Además, reduce la utilidad de los datos, debido a que los ítems incorrectos de la prueba previa no se pueden utilizar en el formulario definitivo de la prueba. Por ello, es importante recurrir a revisores calificados y proporcionarles tiempo suficiente para la revisión y la corrección.

Los cuadernillos de la prueba previa se deben revisar al recibirlos de la imprenta. Se deben realizar comprobaciones aleatorias en cada paquete o caja de cuadernillos para garantizar lo siguiente:

- Todas las páginas se han impreso claramente.
- Las páginas aparecen en el orden correcto.
- No hay páginas duplicadas.
- Los estímulos de lectura de cada unidad están en las páginas correctas.
- Las ilustraciones son claras.

Los cuadernillos de la prueba previa se deben imprimir y comprobar con suficiente antelación, antes de enviarlos a las escuelas. Contar con un plazo suficiente permitirá volver a imprimirlos, si es necesario. Debido a que, por lo general, las tiradas de las pruebas previas son pequeñas, volver a imprimirlas, si es necesario, representa un gasto relativamente pequeño.

IMPLEMENTACIÓN DE LA PRUEBA PREVIA

Los estudiantes no deben tener ninguna duda acerca de cómo responder a las preguntas de la prueba previa o de las fases de la prueba final. Las pruebas se han diseñado para comprobar los conocimientos en un área curricular importante, no la habilidad de los estudiantes para adivinar cómo responder. Los estudiantes deben tener suficientes oportunidades para responder ítems de práctica durante las pruebas previas, tanto al comienzo de la prueba como al comienzo de las secciones dentro de la prueba. Es especialmente importante proporcionar un número suficiente de ítems de práctica (por ejemplo, tres o cuatro) a los estudiantes de sistemas educativos sin tradición en la realización de pruebas del tipo de respuesta múltiple.

El número de ítems de los formularios de la prueba previa puede ser igual o ligeramente inferior al de los formularios finales. Es importante que todos los estudiantes respondan a todos los ítems de la prueba previa. Si la prueba previa es demasiado larga o tiene demasiados ítems difíciles al final, entonces pocos de los últimos ítems de la prueba tendrán respuestas.

RECUADRO 4.1**Ejemplo de pregunta de respuesta múltiple y formato abierto**

$$13 + 7 + 8 =$$

(A) 28 (B) 30 (C) 38 (D) 110

○

$$13 + 7 + 8 = \underline{\hspace{2cm}}$$

Es conveniente que los formularios comiencen con algunos ítems sencillos para animar a los estudiantes menos aventajados a realizar la prueba. A menudo, es recomendable combinar la dificultad de ítems sucesivos para que los estudiantes perseveren en lugar de rendirse al encontrar una serie de ítems difíciles. Intente que la dificultad global de cada prueba previa sea aproximadamente similar. Evite que algún formulario esté lleno de ítems difíciles, porque es posible que los estudiantes renuncien a intentarlo. Si esto ocurre, no habrá datos suficientes de los últimos ítems del formulario para analizar la prueba previa.

Las pruebas previas permiten ensayar versiones alternativas de los ítems en distintos formularios. Las pruebas previas permiten, por ejemplo, ensayar un ítem con formato de respuesta múltiple y el mismo ítem con formato abierto (recuadro 4.1).

También se pueden ensayar distintos textos para un ítem de respuesta abierta. Tenga en cuenta que no se deben utilizar versiones alternativas para los ítems de enlace, estos ítems deben ser idénticos.

CORRECCIÓN DE LA PRUEBA PREVIA

La finalidad de la recogida de datos piloto de las pruebas previas es obtener información que ayude a seleccionar los ítems de buena calidad para la prueba final. En general, no es necesario que las puntuaciones y los nombres de los estudiantes estén vinculados. Los principales

problemas para la corrección de las pruebas previas son el control de la calidad y la coherencia en el tratamiento de las respuestas de los estudiantes.

Todas las correcciones requieren procedimientos de control de calidad. Normalmente, el ingreso manual de datos y puntuaciones de pruebas previas suele ser rentable debido a que el número de ítems es manejable. Los evaluadores y el personal de captura de datos deben estar debidamente capacitados. El encargado del desarrollo de la prueba debe velar por que la calidad del trabajo del personal mencionado sea revisada dos veces al día para garantizar la consistencia y confiabilidad. Las comprobaciones pueden ser más frecuentes en las primeras etapas y tal vez menos frecuentes en etapas posteriores, si se observa que el trabajo de un evaluador es satisfactorio.

Conocer el porcentaje de estudiantes que no responden a los ítems de una prueba previa proporciona información útil sobre cómo estructurar el formulario definitivo de la prueba. A continuación se proporcionan pautas generales relativas a esta cuestión:

- Si al menos el 15 % de los estudiantes no respondieron a varios ítems en la parte final de la prueba, es posible que la prueba previa haya sido demasiado larga. Considere hacer la prueba final más fácil, más corta, o ambas cosas.
- Si al menos el 15 % de los estudiantes no respondieron a un ítem que no estaba en la parte final de la prueba, es posible que haya algún problema en la manera en que se presentó el ítem o que sea demasiado difícil. Posiblemente, los estudiantes pasaron por alto el ítem, no sabían cómo responder o no entendían el texto. Considere revisar y ensayar un ítem nuevo.
- Si un determinado grupo de la población (por ejemplo, el 15 % o más de las niñas) no respondió a una pregunta pero la mayoría de los demás sí lo hizo, es posible que la pregunta esté sesgada. Considere la posibilidad de no incluirla en la prueba final.
- Si al menos el 15 % de los estudiantes sistemáticamente no intentaron responder a ítems con un formato en particular (por ejemplo, ítems de respuesta abierta), es posible que estos estudiantes no entendieran cómo responder o que necesitaran oportunidades para

aprender a responder ítems de este tipo. Considere añadir ítems de práctica o ejemplo adicionales con este formato y ensaye estos ítems de nuevo.

En general, la puntuación de las preguntas omitidas es 9. Asegúrese de que ningún ítem de la prueba tenga una posible puntuación correcta de 9. Si es así, se debe utilizar la letra X (u otra letra del alfabeto) para indicar la puntuación de las omisiones.

Los evaluadores y correctores deben tener claras las reglas para calificar las respuestas omitidas. Generalmente, una respuesta omitida es aquella en la que el estudiante no ha escrito nada en absoluto. Cualquier intento de responder a un ítem, incluso si es ilegible o ininteligible, se trata generalmente como respuesta incorrecta en lugar de respuesta omitida.

Se deben diseñar guías de corrección de ítems de respuesta múltiple para que el desarrollador o el revisor de la prueba obtengan la mayor cantidad posible de datos útiles de cada ítem.

Por ejemplo, se puede codificar una pregunta con cuatro respuestas posibles como 1, 2, 3, 4, 8 o 9. Se pueden utilizar más números para reflejar más opciones.

Los números 1, 2, 3 y 4 indican la opción seleccionada por el estudiante. Se puede utilizar el valor 7 para indicar que el estudiante ha seleccionado dos o más opciones y que, posiblemente, no comprende cómo responder a ítems de respuesta múltiple. En el libro 4 de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*, utilizamos un valor de 8 para indicar que el estudiante no respondió al ítem y de 9 para indicar que un ítem en particular no se planteó al estudiante (ya que estaba en otro formulario de prueba) y, por tanto, no se debe calificar como incorrecto.

Los ítems de respuesta múltiple nunca se deben ingresar en una computadora como correctos o incorrectos. El diseño de los ítems de respuesta múltiple debe seguir un patrón implícito de numeración, desde el 1 al 4 o el 5, en función del número de opciones.

El corrector o la persona responsable de la captura de datos simplemente registra el número (implícito) de la opción seleccionada por el estudiante para cada pregunta de respuesta múltiple. El corrector o la persona responsable de la captura de datos no necesita saber cuál es la

opción correcta o si la respuesta del estudiante es correcta o incorrecta. Una hoja de entrada puede tener el aspecto del ejemplo del recuadro 4.2.

El redactor de los ítems debe proporcionar al analista de datos una lista de las opciones correctas o claves para cada ítem, que el analista ingresará en el programa informático. A continuación, el software de análisis registrará las respuestas de los estudiantes como correctas o incorrectas, de acuerdo con la lista de claves. Conocer las opciones incorrectas seleccionadas por los estudiantes proporciona a los redactores de ítems información importante acerca de la calidad de los ítems de respuesta múltiple y su posible utilidad para el formulario de la prueba final. Por ejemplo, si casi ningún estudiante selecciona una de las dos respuestas incorrectas, es evidente que estas dos opciones no sirven como distractores eficaces.

Los ítems de respuesta cerrada y los ítems de respuesta abierta se puntúan habitualmente con 0 (incorrecto), 1 (correcto) o 9 (omitido). Un ítem de crédito parcial se puede puntuar con 0, 1, 2 o 9.

Corregir manualmente ítems de pruebas previas requiere una formación y un control de calidad similares a los necesarios para puntuar manualmente la prueba final. Un redactor de ítems experimentado debe impartir la formación y supervisar la corrección de los ítems de la prueba previa. El disco compacto adjunto contiene ejemplos de guías de corrección de ítems de respuesta abierta para lenguaje, matemáticas y ciencias.

RECUADRO 4.2

Ejemplo de hoja de captura de datos para una prueba previa

Alumno	P1	P2	P3	P4	←
Ahmed Buta	2	3	2	1	}← Opción seleccionada por cada alumno para cada pregunta
Miriam Wisim	4	3	2	4	
Almet Duras	2	3	1	4	

Los redactores de ítems utilizan las respuestas de las pruebas previas para revisar y perfeccionar sus guías y categorías de corrección antes de comenzar la corrección de la prueba previa. Antes de que comience la corrección manual, los redactores de ítems deben tomar una muestra de cuadernillos de la prueba previa completados y comparar las respuestas reales de los estudiantes a ítems de respuesta corta con las respuestas previstas en las guías de corrección. Los redactores de ítems deben utilizar las muestras para incluir ejemplos de respuestas de estudiantes en sus guías de corrección. Las guías de corrección deben incluir tanto respuestas incorrectas como correctas. La guía de corrección de ítems de crédito parcial del recuadro 3.12 muestra ejemplos de respuestas reales de estudiantes para cada categoría de la guía de corrección, incluidas respuestas de valor cero.

Los redactores de ítems deben perfeccionar o ampliar sus guías de corrección para tener en cuenta el rango de respuestas reales proporcionadas por los estudiantes. En ocasiones, estas revisiones pueden ser muy extensas. Los estudiantes tienden a dar respuestas imprevistas aunque correctas, o formas inusuales pero precisas de expresar sus ideas. Se deben añadir ejemplos así a las guías de corrección, si son razonablemente habituales. Si muchos estudiantes dan respuestas difíciles de clasificar como correctas o incorrectas, los redactores de ítems tienen que mejorar sus guías de corrección para clarificar estas distinciones. Una comisión, o el encargado del desarrollo de la prueba, debe revisar las guías de corrección actualizadas antes de corregir los cuadernillos de la prueba previa.

Durante la corrección manual, los redactores de ítems deben recibir información de los correctores acerca de mejoras adicionales que pueda ser conveniente realizar en las guías de corrección. Si es necesario revisar de forma sustancial la guía de corrección para un ítem, se tendrá que cambiar la puntuación del ítem en función de la guía de corrección revisada, para garantizar la coherencia.

Es esencial revisar las guías de corrección para que los criterios de corrección y los ejemplos proporcionados reflejen el rango real de respuestas de los estudiantes. Si las guías de corrección no se revisan, se perderán algunos ítems debido a la posibilidad de que ninguna de las respuestas de los estudiantes cumpla los requisitos demasiado exigentes de la guía. La corrección de otros ítems puede ser poco

confiable debido a que los evaluadores no sepan cómo corregir las respuestas si no entran dentro de las pautas de calificación y tengan que aplicar su propio criterio personal.

Por lo general, las puntuaciones más altas en ítems de crédito parcial denotan una respuesta más sofisticada o extensa. Una puntuación de 2 sugiere una respuesta «mejor» que una puntuación de 1. Los datos de ítems de crédito parcial de las pruebas previas se pueden utilizar para recopilar información sobre las categorías de respuestas de los estudiantes, la cual puede ayudar a perfeccionar los ítems de la prueba previa o las guías de corrección. Las respuestas a un ítem de crédito parcial de una prueba previa se pueden codificar como 0, 1, 2 o 3, aunque puede que estas puntuaciones no sean jerárquicas. En sistemas de puntuación no jerárquicos, una puntuación de 3 no se considera más sofisticada que una puntuación de 2 o 1. Cada una de las puntuaciones—1, 2, y 3—denota una respuesta correcta, pero de tipo diferente. Por ejemplo, puede haber tres maneras diferentes de resolver un problema de matemáticas. La guía de corrección puede ser bastante compleja y permitir estas tres posibilidades. Si todos los estudiantes de la prueba previa eligen el mismo método, el redactor de ítems puede revisar la guía de corrección para centrarse en el método más popular con una breve referencia a otras posibilidades. La guía de corrección se revisaría para la prueba final para incluir una puntuación de 1 para una respuesta correcta, independientemente del método utilizado para resolver el problema.

Los redactores de ítems deben informar al analista de datos cuando se utilicen ítems de crédito parcial para denotar categorías en lugar de jerarquías, lo que permitirá al analista diferenciar entre las respuestas. El analista puede asignar una puntuación de 1 a cada respuesta de la categoría correcta. Por lo tanto, es muy importante que los correctores comprendan cuándo están corrigiendo ítems *jerárquicos* de crédito parcial y cuándo *categorías* de crédito parcial. Todos los ítems de crédito parcial de las pruebas finales se deben tratar como jerárquicos.

El volumen 4 de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*, incluye una sección especial sobre el análisis de los datos de una prueba previa o piloto. Expone los enfoques de análisis tanto de la teoría clásica de las pruebas (TCP) como de la teoría de respuesta al ítem (TRI). La TRI se utiliza con

frecuencia para analizar ítems de pruebas, vincular formularios de prueba y desarrollar escalas para reportar los resultados en evaluaciones nacionales (Beaton y Johnson 1989). Tiene una serie de ventajas cuando se aplica al cambio de escala de los datos de evaluaciones. La TRI permite caracterizar un ítem de forma independiente de cualquier muestra de individuos que responden al mismo y permite caracterizar a un encuestado individual independientemente de cualquier muestra de ítems planteados a esa persona. Por tanto, la TRI es particularmente útil cuando se plantean varios conjuntos de ítems a estudiantes en una evaluación. Sin embargo, también tiene algunas desventajas—en particular, la complejidad del procedimiento, que requiere considerable habilidad y experiencia—. Cuando en un país no se dispone de la habilidad y experiencia necesarias, se considera aceptable la aplicación de la TCP.

CONFIABILIDAD

Tanto las pruebas previas como las pruebas finales deben proporcionar evidencias acerca de la confiabilidad de las mismas. Una medida de la confiabilidad es un indicador de la consistencia de los resultados de la prueba. La confiabilidad depende de la calidad de los ítems de una prueba, la prueba misma, la forma en que se administran las pruebas, las características del grupo de alumnos (por ejemplo, el esfuerzo que hacen al realizar la prueba previa o las pruebas de evaluación nacionales) y la calidad de la corrección de los ítems. La confiabilidad se trata en el volumen 4, *Análisis de los datos de una evaluación nacional del rendimiento académico*.

Los índices de confiabilidad de la prueba varían entre 0 y 1, donde 0 representa una prueba en la que las respuestas de los estudiantes son totalmente inconsistentes (por ejemplo, una prueba en la que todos los estudiantes responden todos los ítems al azar) y 1 representa una prueba que mide un dominio con coherencia perfecta.

La agencia implementadora debe obtener evidencias de confiabilidad acerca de la medida en que los ítems individuales de la prueba previa se relacionan entre sí. Esta información proporcionará una medida de la coherencia interna de los ítems de la prueba.

Tenga en cuenta que este enfoque presupone que los ítems seleccionados miden un único constructo o rasgo, como la capacidad para el lenguaje o las matemáticas. Normalmente, los equipos de las evaluaciones nacionales e internacionales tienden a omitir los ítems que no son relativamente homogéneos, es decir, que no miden un único constructo o rasgo. La homogeneidad se puede evaluar aplicando un enfoque como el alfa de Cronbach, las fórmulas 20 o 21 de Kuder-Richardson, o un coeficiente de confiabilidad dividido en dos mitades, todo lo cual se puede encontrar en el software estadístico SPSS®.

Si las pruebas de la evaluación incluyen ítems de respuesta abierta o de respuesta libre, la agencia implementadora debe comprobar que el método de corrección sea fiable. Este organismo debe asegurarse de que todos los evaluadores o correctores de ítems de respuesta abierta estén capacitados para juzgar si las respuestas de los estudiantes son aceptables. La formación para ello requerirá que los correctores trabajen con el equipo de desarrollo de las pruebas para elaborar una lista de respuestas aceptables y no aceptables para cada pregunta de respuesta abierta. Después de la formación, los correctores, trabajando en parejas independientes, deben corregir todos los ítems de respuesta abierta de al menos sesenta cuadernillos de la prueba previa seleccionados aleatoriamente. A continuación, se debe calcular el porcentaje de coincidencia exacta entre los evaluadores para el conjunto completo de ítems. Cuando sea evidente que hay confusión entre los correctores acerca de si una respuesta determinada es aceptable, la agencia implementadora debe pedir a los desarrolladores de la prueba que proporcionen aclaraciones. La corrección cuidadosa de los ítems de respuesta abierta debe contribuir a garantizar que apenas haya espacio para desacuerdos acerca de las respuestas aceptables o no aceptables a la hora de corregir los ítems de una evaluación nacional. El disco compacto que se adjunta contiene una serie de ejemplos de correcciones de ítems de respuesta abierta.

SELECCIÓN DE LOS ÍTEMS DE LA PRUEBA

La selección de ítems de la prueba previa para la prueba final (tratada en cierto detalle en *Análisis de los datos de una evaluación nacional del rendimiento académico*, volumen 4 de esta serie), depende en primer lugar, y de manera clave, del marco de evaluación, especialmente de las especificaciones técnicas. En segundo lugar, depende de las propiedades de medición de los ítems.

Normalmente, se adoptan los siguientes criterios de selección para cada ítem:

- El ítem se ajusta a las especificaciones técnicas.
- El porcentaje de alumnos que responde correctamente al ítem oscila entre el 40 y el 80 por ciento.
- El ítem tiene una tasa baja de respuestas faltantes.
- El índice de discriminación (correlación entre el puntaje del ítem y el puntaje total de la prueba) es superior a 0,2.
- La inclusión de este ítem mejora la confiabilidad de la prueba.
- El sesgo del ítem está dentro de límites aceptables para los grupos de alumnos principales.

Las siguientes consideraciones son específicas para los ítems de respuesta múltiple:

- La correlación punto biserial de la clave es positiva y superior a 0,2.
- Todos los distractores son verosímiles (es decir, han sido seleccionados por al menos el cinco por ciento de los alumnos) y tienen una correlación punto biserial cero o negativa.

La tabla 5.1 muestra un resultado típico del análisis de un ítem de respuesta múltiple. A nivel estadístico, el ítem funciona bien.

Los encabezados muestran el número de categorías u opciones del ítem (A, B, C, D). La opción D es la clave u opción correcta y esto se muestra con una puntuación de 1 entre corchetes. Las opciones A, B y C se muestran con una puntuación de 0 entre corchetes. La fila rotulada como “Recuento” muestra el número de alumnos que seleccionó cada opción: 254 alumnos seleccionaron la opción correcta. La fila rotulada “Porcentaje” presenta datos porcentuales (el recuento expresado como porcentaje del número de alumnos). El 67 por ciento de los alumnos eligió la opción correcta. Este resultado muestra que el ítem está dentro de un rango de dificultad aceptable. El ítem es relativamente fácil. Sólo el 3,7 por ciento de los alumnos seleccionó la opción B, lo que sugiere que esta opción es débil o inverosímil. Reescribir esta opción para hacerla más verosímil podría mejorar este ítem y el ítem tendría que volver a ser sometido a una prueba previa. La fila siguiente muestra la correlación punto biserial para cada opción. La correlación punto biserial para la respuesta correcta es de 0,39. Las correlaciones punto biserial para las respuestas incorrectas son todas negativas.

TABLA 5.1

Ejemplo de resultado del análisis de un ítem de respuesta múltiple

Criterio	Opción			
	A [0]	B [0]	C [0]	D [1]
Recuento	90	14	21	254
Porcentaje	23,7	3,7	5,5	67,0
Correlación punto biserial	-0,26	-0,21	-0,16	0,39
Capacidad media	-0,02	-0,48	-0,14	0,54

En los ítems de respuesta múltiple, la correlación punto biserial de la clave es igual al índice de discriminación del ítem. La última fila muestra la capacidad media. La capacidad media de los alumnos que eligieron la opción correcta es considerablemente más alta que la de aquellos que seleccionaron opciones incorrectas. Este resultado también muestra que el ítem funciona bien.

Las siguientes consideraciones son específicas para los ítems de respuesta abierta:

- Si el ítem se califica de manera dicotómica, el índice de discriminación (la correlación entre el puntaje del ítem y el puntaje total) es superior a 0,2.
- Si el ítem otorga crédito parcial, el índice de discriminación es positivo y superior a alrededor de 0,3.
- Si el ítem se califica de manera dicotómica, la mayoría de los ítems deberían estar en el rango de dificultad del 40 al 80 por ciento.
- Si el ítem otorga crédito parcial, cada categoría de puntaje debe recibir al menos el cinco por ciento de las respuestas.
- Si el ítem otorga crédito parcial, el porcentaje general de alumnos que responde a un ítem de manera correcta, calculado combinando las respuestas a diferentes categorías parcialmente correctas, está en el rango del 40 al 80 por ciento.
- Si el ítem otorga crédito parcial, la capacidad media de los alumnos se reduce claramente desde las categorías más altas a las categorías más bajas de ítems de crédito parcial.

En la tabla 5.2 se muestra un resultado típico del análisis de un ítem de respuesta abierta y de crédito parcial con buenos valores estadísticos. Los encabezados de las columnas muestran las categorías de las respuestas de los alumnos. Los alumnos obtuvieron un puntaje de 0 si respondieron al ítem de manera incorrecta. Una respuesta parcialmente correcta obtuvo un puntaje de 1 y una respuesta completamente correcta obtuvo un puntaje de 2. Las respuestas faltantes aparecen como 9 y también obtuvieron un puntaje de 0. El índice de discriminación general es de 0,47, que es alto. Nótese que para los ítems de crédito parcial, el índice de discriminación no es igual a la correlación punto biserial de la categoría con mayor puntaje. El recuento y el porcentaje de respuestas correctas figuran en las dos

TABLA 5.2**Ejemplo de resultado del análisis de un ítem de respuesta abierta y de crédito parcial**

Criterio	Categoría de respuesta de los alumnos				Índice de discriminación = 0,47
	0 [0]	1 [1]	2 [2]	9 [0]	
Recuento	1466	425	268	809	
Porcentaje	49,4	14,3	9,0	27,3	
Punto biserial	0,09	0,11	0,45	-0,48	
Capacidad media	-1,66	0,53	0,90	-1,90	

primeras filas. Casi la mitad de los alumnos que respondieron a este ítem lo hicieron incorrectamente. Más del cinco por ciento seleccionó cada una de las categorías de crédito parcial, lo que sugiere que vale la pena conservarlas. La correlación punto biserial crece desde el puntaje de 0 hasta el puntaje de 2, mostrando que las categorías se comportan de acuerdo a lo esperado. La capacidad media de los alumnos que obtuvieron puntajes de categoría 2 es $-0,9$. Los alumnos que obtuvieron puntajes de categoría 0 tienen una capacidad media de $-1,53$. La diferencia es superior a 0,5 y respalda el mantener las dos categorías en las guías de corrección porque diferencian entre estudiantes de capacidades bastante diferentes.

El porcentaje de alumnos que no respondió a este ítem es bastante alto: el 27,3 por ciento. Esta tabla debe ser analizada dentro del contexto de respuestas faltantes de la prueba en general. En este caso, la mayor parte de los ítems de respuesta abierta tuvieron porcentajes de respuestas faltantes de más de 20. El problema fue causado por la falta de familiaridad de los alumnos y la reticencia a responder a este tipo de ítems, más que por un problema particular con el ítem.

El nivel de dificultad general de la prueba final debería ser adecuado para su propósito. Una prueba final diseñada para monitorear el desempeño de todos los alumnos en la población objetivo debería tener un nivel de dificultad que se ajuste a la capacidad de la población. Las evaluaciones diseñadas para diferentes propósitos, tales como identificar a los estudiantes que alcanzan un valor de referencia

predeterminado, pueden incluir muchos ítems fáciles o muchos ítems difíciles, dependiendo de cuál sea esa referencia.

La experiencia hasta la fecha en el desarrollo de pruebas previas para evaluación nacional sugiere que los redactores de ítems tienden a desarrollar ítems que, en conjunto, son demasiado difíciles. Parte de esta tendencia puede provenir de la experiencia previa de estos redactores elaborando preguntas para exámenes públicos, en los cuales las preguntas tienden a tener un grado bastante alto de dificultad. Más aún, los redactores de ítems de pruebas previas tienden a vivir en zonas urbanas y no tienen una apreciación correcta de los bajos niveles de rendimiento que se pueden encontrar en zonas rurales remotas.

Si muchos de los ítems de prueba previa son demasiado difíciles y si la prueba previa no tiene suficientes ítems fáciles como para ajustarse a los criterios establecidos en las especificaciones técnicas, se requerirá otra ronda de pruebas previas con una nueva serie de ítems más fáciles. Asimismo, si no hay suficientes ítems difíciles, se necesitarán pruebas previas adicionales con ítems más complejos.

La persona que analizará los datos de la prueba final también debería analizar los datos de la prueba previa. Cualquier problema con la forma en que los redactores de ítems informan acerca de las clasificaciones de ítems y las claves, o con la forma en que se otorgó puntaje a los ítems, puede resolverse durante el análisis de la prueba previa.

El equipo de redacción de ítems debería participar de la interpretación del análisis de datos de la prueba previa y de las decisiones respecto de qué ítems retirar de la prueba final y qué ítems con valores estadísticos débiles incluir de todas maneras. Las decisiones respecto de la inclusión o exclusión de ítems deben tener en cuenta las especificaciones técnicas y las áreas del plan de estudios que la prueba debe evaluar.

En general, solo se utiliza un modelo de prueba final por grado. Puede consistir en dos o tres pruebas separadas, tales como una prueba de matemáticas, una de lectura y una de escritura. Estas pruebas pueden combinarse en un cuadernillo único, o imprimirse en cuadernillos separados.

Los problemas de las pruebas previas relacionados con la necesidad de contar con ítems de enlace y su selección y ubicación también se aplican a la prueba final, de existir varios modelos.

Algunos ítems tienen mucho sentido conceptual pero presentan valores estadísticos deficientes. Este resultado puede indicar un problema con la forma en la que el ítem fue presentado. Los alumnos pueden no estar familiarizados con el vocabulario o la forma en la que deben mostrar sus respuestas, o el material de estímulo puede resultar confuso. Idealmente, los ítems con valores estadísticos muy deficientes deberían ser revisados y ensayados una vez más. Sin embargo, cuando ítems con valores estadísticos deficientes abordan criterios importantes de las especificaciones técnicas y no hay otros ítems disponibles, puede ser necesario incluirlos en la prueba final.

En principio, los ítems nunca deberían ser alterados entre la prueba previa y los modelos finales porque la alteración podría afectar las estadísticas del ítem de forma impredecible. En la práctica, las agencias de pruebas tienden a hacer alteraciones menores sobre algunos pocos ítems, generalmente no más de cuatro o cinco en una prueba de 30 ítems. Estas alteraciones menores pueden incluir:

- Cambiar una o dos palabras para mejorar la claridad o para reducir la dificultad del vocabulario.
- Retirar la opción más débil en un ítem de opción múltiple de cinco opciones.
- Corregir errores gramaticales o mejorar la claridad de expresión.
- Mejorar la disposición, como por ejemplo la posición de las etiquetas en un diagrama o la coherencia de los títulos.

NOTA

1. Los ejemplos que se utilizan en este capítulo están basados en análisis de ítems usando el enfoque de la teoría clásica de las pruebas. El libro 4 de esta serie, *Análisis de los datos de una evaluación nacional del rendimiento académico*, cubre este enfoque en más detalle. También presenta otro método de análisis de ítems, la teoría de respuesta al ítem, que utiliza un enfoque estadístico y una terminología diferentes.

PRODUCCIÓN DE LA PRUEBA DEFINITIVA

DISEÑO DE LA PRUEBA DEFINITIVA

El analista de datos o estadístico debe participar en el diseño del formulario definitivo. Debe verificar que se cumplan los siguientes requisitos:

- La organización de la información sobre el alumno en la portada de la prueba es adecuada para el análisis.
- El método para registrar las respuestas a los ítems es adecuado para el análisis.
- Las propiedades y el alcance de las guías de corrección son adecuados para el análisis.
- Las vinculaciones horizontales con datos sobre pruebas de años anteriores o las vinculaciones verticales que pudieran requerirse son confiables desde el punto de vista estadístico.

La información sobre el alumno que debe figurar en la portada de los cuadernillos de prueba está relacionada con el objetivo de la prueba y con el modo en que los datos de la prueba serán reportados. Los alumnos deben ser capaces de brindar la información solicitada

de manera fácil y precisa. La información de la portada normalmente incluye lo siguiente:

- Nombre de la escuela
- Nombre completo del alumno
- Sexo
- Edad o fecha de nacimiento
- Grado o clase a la que concurre
- Información sobre el idioma

En algunos países, los alumnos suelen contar con un número único de identificación nacional. En caso de estar disponibles, estos números deben utilizarse.

También es útil incluir un apartado en la portada para que el examinador registre si los alumnos no realizaron parte o la totalidad de la prueba por ausencia o enfermedad, o si se brindó ayuda especial para escribir las respuestas a los alumnos con discapacidades (véase el recuadro 6.1).

Generalmente, es más fácil garantizar que la identidad de los alumnos esté bien registrada y vinculada con los datos correspondientes si se agrupan todas las pruebas en un solo cuadernillo. Se pueden evitar posibles problemas de identificación si la información de la portada del cuadernillo de pruebas único se ingresa de manera legible y precisa antes de la primera sesión de evaluaciones. Si se utiliza un cuadernillo único, el examinador debe tomar los recaudos necesarios para asegurarse de que estos sean entregados a los alumnos correspondientes en cada sesión de prueba subsiguiente.

Es necesario contar con procedimientos eficaces para que los candidatos reciban los cuadernillos correspondientes cuando se utilizan varios cuadernillos de pruebas. Cuando se utilizan varios cuadernillos, se presentan los siguientes riesgos:

- Los alumnos pueden escribir sus nombres de manera diferente en cuadernillos diferentes.
- Los alumnos pueden usar nombres diferentes en cuadernillos diferentes; por ejemplo, una forma abreviada de su nombre en un cuadernillo y su nombre completo en los otros, un nombre religioso o cultural en un cuadernillo y el apellido en otro, o su primer nombre en un cuadernillo y su segundo nombre en otro.

RECUADRO 6.1**Ejemplo de una portada de pruebas**

EVALUACIÓN DEL MONITOREO DE LA EDUCACIÓN: CUARTO GRADO

Esta sección debe ser completada por el alumno:

Escuela _____

Provincia _____

Grado _____

Nombre _____

Apellido _____

Soy niño Soy niña Edad: años y mesesEl idioma que más hablo en mi casa es el inglés. Sí No **Esta sección debe ser completada por el examinador:**

El alumno estuvo ausente en las siguientes pruebas de este cuadernillo:

Lectura Matemáticas El alumno recibió ayuda especial en lectura matemáticas

Describir la ayuda brindada: _____

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

- Los alumnos pueden escribir su nombre completo, o parte de él, de manera ilegible en, al menos, uno de los cuadernillos.
- Los alumnos pueden omitir escribir su nombre en uno o más cuadernillos.

La organización de la información y las guías generales sobre la administración de las pruebas deben ser claras y concisas. En lo posible, el diseño de la prueba previa y el de la prueba definitiva deben ser iguales.

La prueba debe contener ítems fáciles al principio para dar confianza a los alumnos más flojos. Luego, deben continuar los ítems que presenten algún grado de dificultad, sin seguir ningún orden en particular, de manera que los alumnos no decidan abandonar tras

esforzarse por resolver una serie de ítems difíciles. También es importante brindar a los alumnos más lentos la oportunidad de intentar algunos ítems difíciles, ubicando dichos ítems en la primera parte de la prueba. La prueba debe finalizar con ítems difíciles, pues los alumnos con menor capacidad son menos propensos a completarla. Los ítems relacionados con un estímulo en común (por ejemplo, un párrafo o un mapa) deben estar juntos, sin importar su nivel de dificultad.

Las guías generales sobre la ubicación de los ítems de enlace son las mismas para el formulario definitivo que para la prueba previa (véase el capítulo 4). Los ítems de vinculación horizontal se utilizan para crear vínculos con pruebas anteriores cuando se compara el desempeño en el tiempo. Los ítems de vinculación vertical se necesitan para comparar el desempeño entre grados escolares. Los ítems de enlace deben ubicarse (a) al principio o en la parte central del cuadernillo y (b) en ubicaciones similares en cada cuadernillo de prueba.

Los ítems en los cuadernillos de prueba deben estar identificados en escala de grises para poder ubicarlos (véase el capítulo 3).

El encargado del desarrollo de las pruebas debe entregarle una planilla al analista de datos en la cual se indique dónde aparecen los ítems en cada cuadernillo, incluyendo los ítems de enlace.

La manera en que los alumnos anotarán sus respuestas debe haber sido definida durante el diseño de las especificaciones técnicas (véase el capítulo 2). Generalmente, los alumnos anotan la respuesta al ítem en los cuadernillos de prueba. El diseño de los ítems debe brindar espacio suficiente para que los alumnos anoten sus respuestas y también debe proporcionar un espacio para que los evaluadores anoten su puntuación.

En otras instancias, especialmente en los grados de nivel superior, es posible utilizar formularios de respuestas por separado. Estos formularios deben mostrar claramente cómo vincular la respuesta al ítem o la opción impresa en el cuadernillo de prueba con su ubicación en el formulario de respuestas. Si los ítems en el cuadernillo de prueba están organizados en unidades, lo más práctico es organizar los formularios de respuestas en unidades similares.

IMPRESIÓN Y CORRECCIÓN

Los siguientes factores determinan la extensión del cuadernillo de prueba:

- La cantidad de áreas temáticas cubiertas
- La amplitud de la cobertura dentro de las áreas temáticas
- El formato de los ítems
- Las ilustraciones
- La extensión del material de estímulo
- El tamaño de la fuente
- El presupuesto para la impresión

Los redactores de los ítems deben conocer desde el principio la cantidad de páginas sugeridas para el formulario definitivo de la prueba. Si es evidente que la extensión de la prueba será limitada, tanto el material de estímulo como los diagramas e ilustraciones deberán reducirse. Independientemente del espacio disponible, el diseño de los ítems debe ser claro y ordenado.

Los cuadernillos con menos de 20 páginas (10 hojas) normalmente se imprimen en hojas tamaño A4 (210 × 297 milímetros) y se engrapan en un lado. Los cuadernillos más extensos suelen estar impresos en hojas de tamaño A3 (420 × 297 milímetros) y engrapados en el centro.

Generalmente, los cuadernillos que contienen más páginas permiten incluir una amplia variedad de material de estímulo interesante y también ítems más creativos. El inconveniente es que imprimirlos y distribuirlos es más costoso. También ocupan más lugar durante el embalaje y la corrección, lo que puede aumentar considerablemente los costos totales.

Normalmente, las pruebas se imprimen en ambas caras de la hoja. La calidad del papel debe ser adecuada para asegurar que los ítems impresos de un lado de la hoja no interfieran en la lectura de los ítems impresos del otro lado de la misma hoja.

Las fotografías requieren papel de alta calidad para lograr mayor nitidez. Los diagramas con muchos detalles también deben ser impresos en papel de alta calidad.

Si los alumnos deben responder en los cuadernillos de prueba, el papel de los mismos debe ser suficientemente fuerte para que escriban sus respuestas sin romperlo y para que escriban en ambas caras de la hoja sin que se vea la respuesta del otro lado.

El papel de la portada suele ser de mejor calidad que el que se utiliza para el resto del cuadernillo de prueba, pero esto suele aumentar los costos y no es necesario.

Normalmente, es más práctico imprimir los cuadernillos para diferentes grados con tintas de diferentes colores. Esto ayuda a garantizar que a cada alumno se le entregará el cuadernillo correspondiente. Los colores de tinta seleccionados deben ser fáciles de leer.

Los formularios definitivos deben ser revisados por un corrector con experiencia. También es aconsejable pedir a colegas competentes que no hayan participado en el desarrollo de las pruebas que lean el formulario definitivo para verificar que tenga sentido para quienes usarán la prueba.

Hay dos momentos fundamentales en la revisión de los formularios definitivos de las pruebas. Una vez elaborados los formularios definitivos, deben ser corregidos por el encargado del desarrollo de las pruebas, por los redactores de los ítems correspondientes y, finalmente, por un corrector profesional. Los formularios definitivos deben corregirse por segunda vez cuando se tengan las “pruebas de imprenta”. Las pruebas de imprenta son imágenes de las páginas del cuadernillo tal cual serán reproducidas. La imprenta normalmente entregará las pruebas de imprenta algunos días después de recibir la prueba. El encargado del desarrollo de las pruebas puede aceptar la responsabilidad de corregirlas o puede optar por encargar a un corrector profesional. Se deben destinar al menos dos días para la revisión y corrección de los formularios definitivos. Puede necesitarse más tiempo, según la disponibilidad del personal para la corrección de los cuadernillos.

A menudo, los correctores encuentran cientos de errores pequeños, especialmente, errores en el uso de mayúsculas, puntuación, formato, distribución y ortografía. Si la corrección de las pruebas previas ha sido minuciosa y exhaustiva, y si se han efectuado pocos cambios sustanciales en los ítems, en teoría, la corrección de los formularios definitivos debería detectar pocos errores, o ninguno. En la práctica, esto casi no ocurre. Pueden aparecer errores en los formularios definitivos

donde no aparecieron antes. Los correctores suelen necesitar varios días para revisar exhaustivamente los formularios definitivos de la prueba y las guías generales de administración, aun cuando las pruebas previas han sido corregidas antes.

La impresión de trabajos extensos debe encargarse con varias semanas y hasta meses de anticipación. La imprenta sugerirá una fecha de entrega. En algunas ocasiones, el equipo de evaluación nacional puede negociar el pago de incentivos por la entrega anticipada y proponer descuentos o penalizaciones por el retraso en la entrega. Las imprentas pueden cometer errores en la impresión de los cuadernillos; la falta de algunas páginas en algunos de ellos siendo el error más frecuente. El encargado del desarrollo de las pruebas debe revisar de manera aleatoria las cajas de los cuadernillos definitivos impresos para detectar errores.



CALIFICACIÓN MANUAL DE LOS ÍTEMS

El equipo de corrección nacional debe asegurarse de que los correctores que realizan la calificación manual de los formularios definitivos estén debidamente capacitados. En esta fase, las guías para la evaluación y puntuación de pruebas o rúbricas, revisadas durante la prueba previa, deben ser casi definitivas. Antes de que comience la calificación manual definitiva, los redactores de los ítems pueden seleccionar una pequeña muestra de formularios de prueba completados, verificar la claridad y eficacia de las guías para la evaluación y puntuación, e introducir pequeñas modificaciones.

La instalación de un centro de corrección de pruebas para la calificación manual y el establecimiento de procesos de calificación manual efectivos deben planearse con considerable anticipación. El equipo de corrección nacional debe haber abordado las siguientes preguntas antes de que comience la calificación manual:

- ¿Dónde se guardarán los materiales que conforman las pruebas?
- ¿Cómo serán trasladados a los centros de corrección de pruebas?
- ¿Cómo se garantizará la seguridad de los materiales? (Las pruebas y las guías para su evaluación y puntuación deben custodiarse).

- ¿En qué horarios se realizará la calificación manual? ¿La calificación se organizará en unidades diarias, en turnos (incluyendo un turno nocturno)? ¿Los correctores trabajarán durante los fines de semana?
- ¿Cómo registrarán los datos los correctores?
- ¿Qué herramientas de corrección se necesitan? Generalmente, se prefiere la tinta roja o verde, dado que esos colores se ven claramente en los cuadernillos de los alumnos. Las notas o señaladores autoadhesivos son útiles para marcar aquellos ítems sobre los cuales los correctores tienen dudas.
- ¿Los correctores serán remunerados según la cantidad de pruebas corregidas o según la cantidad de tiempo dedicado a la corrección? Ambos métodos presentan ventajas y desventajas. Si se considera la cantidad de pruebas corregidas, los correctores pueden darse prisa y realizar una corrección descuidada con el fin de acumular la mayor cantidad de pruebas posible. En el caso de la remuneración por tiempo, los correctores pueden no dedicarse completamente a la tarea y el rendimiento puede ser bajo. Un acuerdo equilibrado podría ser la remuneración según el tiempo dedicado, pero con la exigencia de una cantidad mínima de pruebas completadas por día.

En un centro de corrección de pruebas se necesitan correctores, correctores expertos y un jefe de correctores. Los responsables de la selección de los correctores deben entrevistar a los aspirantes y verificar sus referencias.

El jefe de correctores es el responsable de todas las operaciones diarias. Debe garantizar el cumplimiento del horario por parte de los correctores, resolver cuestiones relacionadas con la corrección, supervisar los procedimientos de control de calidad y mantener la seguridad de las pruebas. Además, debe ser un corrector con experiencia y con capacidad demostrada para la gestión de operaciones, preparado para despedir a aquellos miembros del equipo que no realicen su trabajo como se espera.

Los correctores expertos son los responsables de supervisar la corrección de pruebas en un área temática en particular e implementar procedimientos de control de calidad. Cada prueba debe tener al menos un corrector experto (por ejemplo, un corrector experto en

matemáticas y un corrector experto en comprensión lectora). Estos correctores deben ser expertos en un área temática, tener experiencia en la corrección e imponer respeto.

Los correctores asignan puntajes a las respuestas de los alumnos. Normalmente, los maestros son buenos correctores. Deben ser diligentes, constantes y confiables, y deben conocer su área temática.

Generalmente, el encargado del desarrollo de las pruebas designa a redactores de ítems de áreas temáticas pertinentes para que capaciten a los correctores. Preferiblemente, la persona a cargo de la capacitación de los correctores debe ser un experto en la materia. El encargado del desarrollo de las pruebas puede ocuparse de capacitar a los correctores en su área temática. Lo ideal es que la persona que llevó a cabo la capacitación de correctores durante las pruebas previas lo haga también en las pruebas definitivas.

Debe preverse tiempo suficiente para organizar varios periodos de capacitación para cada grupo de correctores. La capacitación debe hacer hincapié en el hecho de que los correctores leerán una variedad de respuestas correctas posibles. Algunas de ellas pueden no ser como las que habitualmente leen, pueden ser diferentes a las respuestas de un libro de texto o pueden estar mal redactadas o contener vocabulario poco convencional. La capacitación de los correctores debe tratar los siguientes aspectos:

- Los correctores tienen poca, o ninguna, libertad para determinar la pertinencia de una respuesta; no hay lugar para opiniones o preferencias personales.
- Los alumnos no deben ser penalizados por las faltas de ortografía o los errores gramaticales en comprensión lectora, matemáticas o ciencia, a menos que el trabajo sea imposible de comprender.
- Los correctores deben consultar a un corrector experto cuando no estén seguros de cómo puntuar una respuesta específica.
- Los correctores deben utilizar la misma puntuación (normalmente 0) de manera uniforme para todas las respuestas incorrectas y todas las respuestas ilegibles o incomprensibles, incluyendo la escritura de una sola letra o una simple línea.
- Los correctores deben utilizar el mismo código de puntuación (normalmente 9) de manera uniforme para indicar que el alumno

no intentó responder el ítem, es decir, que no hay ni siquiera marcas con lápiz en el espacio destinado para la respuesta.

- Para facilitar la captura de datos, los correctores deben utilizar solamente el espacio asignado en el cuadernillo al realizar la calificación manual.
- Los correctores no son los responsables de computar los puntajes para llegar a un total.

Durante la capacitación, debe ponerse el acento en asegurar que los correctores comprendan qué implica la tarea de corregir y asignar puntajes y en mantener la coherencia en la evaluación y puntuación de las pruebas. Los métodos para la capacitación suelen variar. El siguiente método es una sugerencia, pero pueden existir otros.

- El capacitador les pide a los correctores que respondan cada ítem de calificación manual en la prueba. Este proceso ayuda a que los correctores se familiaricen con los ítems y garantiza que los hayan leído y comprendido de manera adecuada.
- El capacitador le entrega a cada corrector cuatro o cinco cuadernillos de pruebas completados. Luego, analiza el primer ítem y la guía de corrección y, finalmente, los correctores asignan una puntuación a este ítem en sus cuadernillos de pruebas. El capacitador fomenta la discusión grupal de dudas o discrepancias sobre cómo asignar puntos a un ítem. Se anima a los correctores a compartir las respuestas que difieran de las respuestas sugeridas en las guías de corrección. Luego de que el primer ítem ha sido debidamente discutido, el capacitador pasa al ítem siguiente y su correspondiente guía de corrección. Este método habitualmente toma varias horas.
- Los correctores participan en una segunda sesión de capacitación en la cual trabajan en pares. Asignan puntuación a algunas pruebas de manera individual y, luego, verifican el trabajo del compañero y discuten las diferencias en sus criterios. Si no logran llegar a un acuerdo, deben consultar al capacitador. Luego de que las sesiones de capacitación han finalizado, el corrector experto asume la responsabilidad de dirigir a los correctores. El corrector experto debe comunicar al capacitador los problemas que puedan surgir durante la corrección de las pruebas.

- Durante la evaluación y puntuación real de los cuadernillos de pruebas, el corrector experto debe seleccionar algunos ítems problemáticos cada día y fomentar pequeñas discusiones sobre ellos para mantener el foco y la coherencia.

Para lograr una buena verificación de los ítems de respuesta corta en la evaluación nacional se debe incluir una revisión inicial de prácticamente todos los cuadernillos de pruebas. Generalmente, esta revisión la realizan los correctores expertos. El tiempo de revisión puede reducirse gradualmente al 10 % o el 20 % de los cuadernillos a medida que los correctores muestran coherencia y confiabilidad en sus evaluaciones y puntuaciones.

Si se emplea un grupo numeroso de correctores, se necesitarán varios correctores expertos para asegurar la calidad de la verificación y para brindar observaciones inmediatas a los correctores sobre los errores que cometan. El corrector experto puede solicitar a los correctores que vuelvan a asignar puntuaciones a algunos ítems en cuadernillos que ya han sido evaluados cuando hayan cometido errores en su corrección. Los procedimientos para la evaluación de discrepancias también deben ser claros. Generalmente, la puntuación asignada por el corrector experto es la que tiene validez.

Las pruebas de lengua pueden incluir uno o más ítems que solicitan una redacción como respuesta. Las redacciones habitualmente son evaluadas dos veces. El segundo corrector asigna un puntaje a la redacción sin conocer el puntaje o la calificación que le ha asignado el primer corrector. Luego, se comparan las dos puntuaciones. Generalmente, se aceptan diferencias de un punto y, en tal caso, los dos puntajes se promedian. En caso de que haya diferencias notables entre los puntajes de los correctores, al menos uno de los dos puntajes deberá modificarse. Este cambio deberá ser consensuado entre ambos correctores. Si los correctores no logran llegar a un acuerdo, el asunto debe enviarse al corrector experto para que decida.

La calificación manual requiere mucha concentración. Los correctores no deben trabajar durante mucho tiempo en un mismo día o sin descanso. Habitualmente, un periodo de entre seis horas y seis horas y media por día es lo máximo que se permite. Un día de trabajo debe

consistir en una sesión de corrección de tres horas en la mañana, un pequeño receso y una sesión de tres horas en la tarde con un pequeño receso. Los correctores más lentos pueden necesitar media hora más para completar el número de pruebas exigido por día. Los correctores deben completar planillas de asistencia diaria.

PARTE

2

ELABORACIÓN DE CUESTIONARIOS

DISEÑO DE CUESTIONARIOS



Un cuestionario es un conjunto de ítems diseñado para obtener información de una persona. La clase de información puede variar ampliamente y puede incluir datos sobre características personales, calificaciones y prácticas laborales, condiciones y recursos laborales o información en cuanto al origen de las personas y sus actitudes, creencias u opiniones sobre ciertos temas.

Una evaluación nacional busca obtener un cálculo estimativo confiable del rendimiento de los estudiantes (medido por medio de un examen diseñado especialmente) e información (medida por medio de un cuestionario) sobre las variables principales asociadas con las diferencias en el rendimiento de los estudiantes. Las pruebas recopilan información sobre el desempeño de los estudiantes y los cuestionarios, cuando se los utiliza junto con las pruebas, recopilan datos sobre variables que podrían asociarse con las diferencias en los niveles de desempeño de los estudiantes o ayudar a explicarlas. Por ejemplo, los datos de los cuestionarios pueden indicar que las escuelas que no cuentan con biblioteca se relacionan con un desempeño deficiente o que las escuelas en las que los docentes participan con regularidad en programas de desarrollo profesional se asocian con un rendimiento estudiantil alto. Estos datos indican formas en las que los recursos educativos podrían canalizarse de manera provechosa para mejorar el aprendizaje de los estudiantes.

Un buen cuestionario recopila datos sobre variables para las que los responsables políticos desean información precisa, variables que ellos pueden modificar y sobre las que desean influir, variables que, según la evidencia aportada por las investigaciones, pueden modificar el rendimiento de los estudiantes.

Un error habitual en el diseño de cuestionarios es la recopilación de demasiada información. Generalmente, los responsables políticos están interesados en información sobre solamente una pocas variables clave. Además, incluso si existen buenas razones científicas para recopilar cierta clase de datos, la consideración de las consecuencias políticas y sociales de la recopilación de datos podría indicar que una evaluación nacional no es el mejor mecanismo para ello.

La información, generalmente, puede recopilarse de otras fuentes que no sean cuestionarios en los países que guardan registros confiables y precisos sobre las características de las escuelas, los docentes y los estudiantes. Vale la pena averiguar si los registros gubernamentales son una fuente útil de información, porque el acceso a tales registros puede ser más barato y fácil que la administración de cuestionarios.

El diseño del cuestionario debe describir claramente qué tipo de datos se recopilarán, cómo se analizarán y registrarán, y de qué manera los resultados podrían contribuir a mejorar la educación. Los pasos principales en el diseño del cuestionario son los siguientes:

- Decidir cuál es el propósito de un cuestionario y cómo se utilizarán los datos.
- Desarrollar un plan que especifique los encuestados, las áreas de interés, los tipos de ítems y el protocolo de codificación o puntaje y de administración (que puede ser completado por el entrevistador o autocompletado).
- Redactar los ítems y utilizar grupos (o comités) de expertos en el tema para revisarlos y mejorarlos, y diseñar el formato del formulario de modo que los encuestados puedan usarlo fácilmente y que quienes ingresan los datos puedan procesarlos eficientemente.
- Especificar un plan de análisis de datos para el procesamiento de la información recopilada y la creación de variables e indicadores de medición, para el análisis estadístico posterior.

- Realizar un cuestionario de prueba previa o de campo para establecer la idoneidad de los ítems y las categorías de respuesta.
- Analizar los datos del cuestionario de prueba previa, mejorarlo y producir los cuestionarios finales para su administración.

La tabla 8.1 brinda detalles de los pasos en el desarrollo del cuestionario y de las personas que participan.

Los cuestionarios y las instrucciones para su administración deben prepararse y someterse a un ensayo previo o de campo al mismo tiempo que las pruebas. De esta manera, los proyectos de los cuestionarios deben desarrollarse al mismo tiempo que los proyectos de las pruebas, y los cuestionarios deben escribirse y planificarse al mismo tiempo que se redactan y ensayan los ítems de las pruebas.

CONTENIDO DEL CUESTIONARIO

Un cuestionario debe recopilar información sobre variables clave que podrían ayudar a explicar las diferencias en el desempeño de los estudiantes en una prueba de rendimiento de los estudiantes. Sin embargo, una infinidad de combinaciones de variables pueden afectar el desempeño de los estudiantes. Un cuestionario puede centrarse en solo unas pocas variables.

Los responsables políticos generalmente desean conocer sobre las variables asociadas con los temas educativos en su país, tales como el idioma de instrucción, las desigualdades en la distribución de los recursos educativos o las actitudes hacia la educación de las niñas. Los responsables políticos quizá no sepan qué variables investigar. Es posible que señalen una lista de variables desde su observación y experiencia personal, o variables que piensan que “deben” figurar en un cuestionario. Es necesario reducir dicha lista a unos pocos requerimientos que posiblemente puedan utilizarse para dar forma al contenido del cuestionario.

Puede que los responsables políticos desconozcan que el proceso de análisis y reporte de los datos del cuestionario es costoso y requiere conocimientos técnicos. Debido a que los recursos siempre son limitados, es necesario que los cuestionarios sean concisos y muy pertinentes. La recopilación de datos debe ser técnicamente aceptable si

TABLA 8.1
Partes del desarrollo del cuestionario

Parte	Descripción	Personas que participan
1. Propósito	Esclarecer el propósito y la posible utilización de los datos del cuestionario.	Responsables políticos, grupos de interés clave y encargado del desarrollo de la prueba
2. Proyecto	Diseñar el modelo de cuestionario para especificar los encuestados, las áreas de interés, los tipos de ítems, la codificación y el protocolo de administración.	Encargado del desarrollo de la prueba, expertos, analistas de datos, redactores de ítems experimentados, docentes experimentados, responsables políticos y grupos de interés clave
3. Ítems	Redactar ítems de cuestionarios. Mejorar para obtener claridad y utilidad en los comités de cuestionarios. Revisar los cuestionarios.	Encargado del desarrollo de la prueba y redactores de ítems Encargado del desarrollo de la prueba y redactores de ítems
4. Plan de análisis de datos	Especificar el plan para procesar la información, para crear variables e indicadores de mediciones y para los tipos de análisis.	Encargado del desarrollo de la prueba, responsables políticos y grupos de interés clave Analista de datos y encargado del desarrollo de la prueba
5. Prueba previa	Diseñar, producir y corregir los cuestionarios para la prueba previa. Redactar instrucciones de administración para la prueba previa de cuestionarios e instruir a los examinadores. Realizar la prueba previa de los cuestionarios al mismo tiempo que se realiza el ensayo de las pruebas.	Encargado del desarrollo de la prueba, redactores de ítems, profesionales de diseño y diagramación y correctores Encargado del desarrollo de la prueba y redactores de ítems
6. Cuestionario final	Analizar los datos de la prueba previa del cuestionario. Mejorar las instrucciones de administración y del cuestionario en función de los datos de la prueba previa y las observaciones aportadas por el administrador de la prueba. Producir el formulario final del cuestionario.	Encargado del desarrollo de la prueba, gerente de logística y examinadores Encargado del desarrollo de la prueba y analista de datos Encargado del desarrollo de la prueba, redactores de ítems y analista de datos Encargado del desarrollo de la prueba, redactores de ítems, profesionales de diseño y diagramación, y correctores

estos se van a utilizar para explicar el desempeño de los estudiantes. Los modelos utilizados por otras evaluaciones nacionales pueden proporcionar una orientación aproximada. Cada país tiene sus propias necesidades, sin embargo, y estas deben determinar la naturaleza adecuada de cada cuestionario.

Es posible que el encargado del desarrollo de la prueba o la persona responsable de la producción del cuestionario tenga que orientar a los responsables políticos sobre las variables clave que pueden brindar información útil. Para esto, es posible que tenga que presentar a dichos responsables ejemplos relevantes para ayudarles a considerar cómo podrían utilizar la información recopilada. Esta información ayudará posteriormente a perfeccionar la lista de variables que se abordarán.

Debido a que los cuestionarios se diseñarán para abordar temas que los encuestados probablemente conozcan, los temas para los estudiantes, padres, docentes y directores serán distintos. En la siguiente sección se sugieren algunos temas apropiados para los cuestionarios de cada uno de estos grupos.

Cuestionarios para estudiantes

Los cuestionarios para estudiantes pueden recopilar la siguiente información:

- Sexo, edad e idioma de origen (generalmente todos se recopilan en el frente del cuadernillo del examen).
- Antecedentes educativos, tales como años de escolaridad y períodos de ausencia escolar.
- Oportunidades de asistir a la escuela.
- Expectativas de éxito y actitudes familiares y personales sobre el valor de la escuela.
- Percepción del entorno del aula, como sentido de seguridad, compañerismo de otros estudiantes o apoyo docente.

Cuestionarios para padres

Los cuestionarios para padres pueden recopilar la siguiente información:

- Nacionalidad, sexo e idioma de origen.
- Entorno doméstico, como acceso a libros, escritorios e iluminación.
- Origen familiar, como educación de los padres e idioma que se habla en el hogar.
- Actitudes hacia la educación, como el compromiso de mandar a los hijos a la escuela, la percepción del valor y de la importancia de la educación o la percepción de la calidad de la educación.
- Atención a las tareas para el hogar y recursos de estudio que se brindan en el hogar para los hijos.
- Accesibilidad económica y física a la educación para los hijos.
- Expectativas de rendimiento educativo de los hijos.
- Participación en las escuelas, tales como la participación en el aula o en comisiones.
- Naturaleza de los informes escolares sobre el progreso de los hijos y sus valores.
- Apoyo financiero para escolaridad en forma de pago de los libros de texto y aranceles.

Cuestionarios para docentes

Los cuestionarios para docentes pueden recopilar la siguiente información:

- Sexo y edad.
- Idioma materno.
- Condiciones de enseñanza, como tamaño de la clase, acceso a recursos, porcentaje de alumnos que cuentan con libros de texto, acceso a docentes suplentes en caso de enfermedad y asistencia con alumnos difíciles.
- Experiencia educativa, formación docente y cantidad de años en la escuela.
- Compromiso profesional con el aprendizaje, como acceso a desarrollo profesional e interés en el mismo, interés en la enseñanza y tiempo empleado en la planificación de las clases.
- Disponibilidad de apoyo educativo mediante visitas al aula por parte de los directores, inspectores escolares o supervisores.
- Metodología de la enseñanza, como idioma de instrucción, utilización de las evaluaciones y estilo de enseñanza.

- Satisfacción con las condiciones laborales, como antigüedad, nivel de salario y supervisión.
- Relaciones con la comunidad escolar, como interacción con los padres, participación en comisiones escolares y en los eventos de la comunidad local.
- Distancia entre el domicilio del docente y la escuela.

Cuestionarios para directores

Los cuestionarios para directores pueden recopilar la siguiente información:

- Sexo y edad.
- Formación y experiencia educativa y en gestión.
- Entorno escolar, como calidad de los edificios e instalaciones, así como disponibilidad de recursos.
- Informes escolares, como la fluctuación de la cantidad de estudiantes, el grado de ausentismo estudiantil y docente, y la frecuencia de cambio de escuela de los estudiantes.
- Compromiso profesional con el liderazgo escolar, como el acceso e interés en el desarrollo profesional y el interés en la educación.
- Estilo de liderazgo y utilización del tiempo.
- Satisfacción con las condiciones laborales, como antigüedad, nivel de salario y frecuencia de supervisión.
- Relaciones con la comunidad escolar, como interacciones con los padres y participación en eventos de la comunidad local.

ESPECIFICACIONES TÉCNICAS DEL CUESTIONARIO

Es necesario un documento de especificaciones técnicas que guíe el desarrollo del cuestionario. Este describe las cuestiones normativas clave que serán el eje del cuestionario; identifica a los encuestados; enumera las variables clave que deben ser abordadas y especifica el formato de los ítems, las categorías de respuesta y el protocolo de administración.

El recuadro 8.1 brinda un ejemplo de especificaciones técnicas del cuestionario utilizadas para recopilar información sobre los valores y las

RECUADRO 8.1**Especificaciones técnicas para el cuestionario sobre valores y actitudes****Parte I**

Áreas de interés	Actitudes hacia la escuela	Creencias sobre la vida en Papúa Nueva Guinea	Percepciones de la comunidad local
Cantidad de preguntas	10	15	15
Encuestados	3.º grado 5.º grado 8.º grado	5.º grado 8.º grado	5.º grado 8.º grado
Categorías de respuesta	Sí o no	Sí o no	Sí o no

Parte II

Actitudes hacia la escuela	Creencias sobre la vida en Papúa Nueva Guinea	Percepciones de la comunidad local
Creencias sobre el rendimiento personal, duración prevista de la escolarización y planes personales a futuro	Actitudes hacia la educación: enseñanza en lengua vernácula, educación obligatoria, rol de las escuelas, educación de las niñas y roles de las mujeres	Nivel de cooperación en la comunidad local percibido: apoyo a las escuelas, participación local en eventos de la comunidad e intercambio de recursos
Percepción de amabilidad por parte de los docentes, compañerismo de los estudiantes, hostigamiento y deseos de hacer amigos fuera del pueblo	Actitudes hacia la comunidad: intención personal de permanecer en la comunidad local o motivos para irse de ella	Actitud percibida de la comunidad local hacia las niñas y las mujeres

Parte III

Valores	Creencias sobre la vida en Papúa Nueva Guinea	Percepciones de la comunidad local
Valores sobre la resolución de conflictos	Actitudes hacia la resolución de conflictos y las peleas	Niveles percibidos de empleo positivo en la comunidad local y utilización de medios pacíficos de resolución de problemas
Valores sobre hábitos de higiene personal	Actitudes hacia el alcohol y las drogas	Percepción de problemas causados por el uso de las drogas y el alcohol en la comunidad local

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

actitudes de los estudiantes hacia la escuela y su comunidad local. Las reformas recientes en educación y los nuevos materiales curriculares presentados a las escuelas habían enfatizado el enseñar a los estudiantes a valorar su comunidad local y a adquirir destrezas que los ayuden a contribuir positivamente a la vida de su pueblo cuando sean adultos. Los responsables políticos de Papúa Nueva Guinea deseaban recopilar información sobre las expectativas de los alumnos y las percepciones de estos sobre la escuela y la comunidad. El cuestionario se administró a todos los estudiantes que realizaron las pruebas de evaluación nacional.

ÍTEMS DE LOS CUESTIONARIOS

Al decidir la cantidad de ítems en un cuestionario deben considerarse varias cuestiones, entre ellas la cantidad de tiempo disponible para responder a las preguntas, los recursos disponibles para el análisis y la complejidad del análisis requerido. Es preferible un cuestionario breve y limitado que se analice adecuadamente y brinde información útil antes que uno largo y exhaustivo que nunca se procesa por completo.

La cantidad de ítems necesarios para medir una variable específica depende de la naturaleza de la variable. Algunas variables, tales como el sexo o la edad, se pueden medir directamente. Otras, tales como el nivel socioeconómico, tienden a construirse mediante distintos ítems, como el nivel de educación de los padres, su situación laboral, la ubicación de la vivienda y la propiedad de los bienes. Los datos derivados de una medición directa se denominan *variable bruta*. Una *variable agregada* combina datos de dos o más ítems para representar un constructo. Generalmente, a los responsables políticos les resulta más fácil interpretar los resultados de análisis basados en variables brutas que los resultados basados en variables agregadas.

Las decisiones sobre si es necesaria una variable bruta o agregada para respaldar un constructo deberían basarse en unas buenas prácticas y condiciones de investigación en el país. Tanto las encuestas nacionales como las internacionales han utilizado variables agregadas. En un estudio internacional, por ejemplo, las variables “leer con un fin utilitario” y “leer por placer” se basaban en variables agregadas (consulte la tabla 8.2).

TABLA 8.2

**Funciones de la comprensión lectora en un estudio internacional:
Ponderaciones utilizadas para crear dos nuevas variables, “leer con fines
utilitarios” y “leer por placer”**

Ítem (abreviado)	Ponderaciones	
	Fines utilitarios	Placer
Me ayuda en la escuela.	0,75	
Me ayuda a aprobar los exámenes.	0,74	
Me ayuda con futuras materias escolares.	0,73	
Me ayuda a trabajar mejor.	0,65	
Puedo ir a la universidad.	0,65	
Me ayuda a conseguir un buen trabajo.	0,63	
Mis padres creen que es importante.	0,58	
Lo disfruto.		0,76
Es emocionante.		0,72
Es interesante.		0,71
Es como viajar a otro mundo.		0,68
Me gusta pensar sobre las cosas que leo.		0,54
Me divierte creer que soy un personaje de la historia.		0,53
Me gusta leer solo.		0,53
Me ayuda a relajarme.		0,50

Fuente: datos extraídos de Greaney y Neuman 1990 (tabla 8, se excluyeron las cargas de menos de 0,2).

Las cuestiones específicas de los países son importantes al momento de decidir cuántos ítems se necesitan para medir una variable. Por ejemplo, en un país donde las condiciones de formación docente son casi uniformes y todos los docentes tienen al menos dos o tres años de formación terciaria en instituciones reconocidas, puede ser suficiente una sola variable bruta que mida los años de formación terciaria. En un país donde las condiciones de formación docente varían ampliamente, la calidad de las instituciones educativas es desigual y es posible que muchos docentes hayan recibido capacitación directamente en el puesto de trabajo, sin embargo, puede ser necesario juntar una cantidad de variables brutas para representar un constructo de formación docente que refleje de manera adecuada esta situación. De manera similar, en un país rico, los recursos de estudio en el hogar pueden medirse simplemente con una variable bruta en relación con el acceso a Internet, pero

en un país pobre, los recursos de estudio en el hogar pueden representarse mejor como la suma de variables brutas, entre ellas el acceso a un escritorio, una silla, una lámpara, lápices, papel y libros de texto.

Las decisiones sobre si utilizar una sola variable bruta o una variable agregada para obtener una medición también depende de las creencias sobre la importancia de las posibles variables brutas. Por ejemplo, al medir la experiencia docente, si la calidad de esta varía ampliamente según dónde están empleados los docentes y se cree que la ubicación del trabajo anterior del docente podría incidir también en el desempeño de los estudiantes, entonces debería recopilarse información sobre dónde ha trabajado el docente al igual que sobre el tiempo que se ha desempeñado en la enseñanza. Si existe la creencia de que la cantidad de años de experiencia docente podría incidir en el rendimiento de los estudiantes, sin importar dónde se ha obtenido esta experiencia, entonces una sola variable bruta probablemente sea suficiente.

FORMATO DE LOS ÍTEMS

Los ítems de elección forzada son bastante más fáciles, rápidos y menos costosos de procesar que los ítems de respuesta abierta. Debido a que los ítems de elección forzada brindan una cantidad limitada de categorías de las cuales seleccionar una respuesta, el procesamiento de datos es simplemente cuestión de ingresar la selección del encuestado en una computadora. Por el contrario, las respuestas a ítems de respuesta abierta deben procesarse manualmente antes de que se ingresen en una computadora.

Los datos del cuestionario suelen resumirse para su informe. Por ejemplo, las respuestas a una pregunta sobre el tiempo que los estudiantes tardan en trasladarse a la escuela cada día pueden clasificarse en unos pocos intervalos amplios, tales como “menos de una hora”, “entre una y dos horas” y “más de dos horas”. En una versión abierta de este ítem, algunos estudiantes proporcionarán el tiempo en minutos y otros en horas, otros podrían escribir “mucho tiempo” y otros responderán de manera ilegible. El rango de respuestas será extenso y clasificarlas necesariamente conllevará un elemento de subjetividad, incluso requerirá el tomar decisiones sobre cómo clasificar las respuestas tales como “mucho tiempo”.

Son preferibles los ítems de elección forzada cuando se puede especular con bastante seguridad el rango posible y las diferencias en las categorías de respuestas de la mayoría de los encuestados. Sin embargo, si existe cierta incertidumbre, entonces pueden utilizarse categorías diferenciadas con más precisión que las requeridas, a efectos del informe. Después de ingresados los datos, se puede decidir qué categorías brindan poca información y cuáles pueden combinarse o descartarse (por ejemplo, si nadie las seleccionó).

Puede ser factible el uso de ítems de respuesta abierta si el cuestionario se administra a una pequeña muestra y hay recursos disponibles para clasificar las respuestas manualmente. Realizar pruebas previas o pruebas de campo con ítems de respuesta abierta a veces es útil para brindar información con el fin de generar categorías para una versión de elección forzada de los ítems en la administración final.

IDIOMA DEL CUESTIONARIO

El idioma utilizado en el cuestionario debe ser aquel en el que es más factible que los encuestados puedan leer y escribir con fluidez. Sin embargo, la elección del idioma debe permitir economías de escala. Los cuestionarios se suelen administrar en el mismo idioma que el material de la prueba.

ENCUESTADOS

Dado que siempre se recopila algo de información general sobre el estudiante en la portada del cuadernillo de prueba, la selección de los encuestados para el cuestionario depende de lo que deseen saber los responsables políticos y de la viabilidad de obtener esta información de manera confiable y eficiente.

Los siguientes son algunos de los problemas relacionados con los encuestados:

- Es posible que los alumnos sean demasiado jóvenes para completar un cuestionario de manera confiable o precisa.

- La falta de recursos puede limitar la administración de los cuestionarios a un grupo pequeño, tal como docentes o directores, en lugar de a miles de estudiantes.
- Es posible que muchos padres sean analfabetos o que no sean confiables en cuanto a la devolución del cuestionario.
- Es posible que los docentes y directores no estén motivados a completar un cuestionario largo o puede que tengan miedo de responder honestamente a una pregunta.

Sea cual fuere la decisión sobre los encuestados, la muestra seleccionada para un cuestionario debe ser representativa de la población. Si el cuestionario se administra a los estudiantes, la muestra que se extrajo para la prueba debe responder el cuestionario. Debe consultarse a los expertos en muestras sobre el tamaño necesario de ellas para su administración a docentes, directores y padres.

ADMINISTRACIÓN DEL CUESTIONARIO

Los cuestionarios normalmente se responden por escrito o se administran en una entrevista. Esto último requiere un entrevistador capacitado para realizar las preguntas y anotar las respuestas del entrevistado (quizá mediante la aplicación de los códigos que se encuentran en el formulario del cuestionario).

En evaluaciones a gran escala, la mayoría de los cuestionarios se escriben y administran en grupo para minimizar los costos. Los cuestionarios deben incluir instrucciones sobre cómo responder las preguntas. Las instrucciones podrían incluir los motivos para la recopilación de la información.

La recopilación de los datos del cuestionario puede realizarse bajo la supervisión de un trabajador de campo, quien recoge los formularios después de que los encuestados los hayan respondido.

PLAN DE ANÁLISIS DE LOS DATOS

Un plan de análisis de los datos especifica la clase de información brindada por cada ítem del cuestionario y cómo se utilizará la

información en el análisis. La asistencia estadística de expertos en el diseño del plan aumentará la certeza de que los datos se podrán analizar significativamente y los resultados serán rigurosos y justificables (consulte el volumen 4, *Análisis de los datos de una evaluación nacional del rendimiento académico*).

El plan deberá reflejar lo siguiente:

- Las características de medición de las variables. La manera en que los datos del cuestionario pueden analizarse depende de las características de medición de las variables. Las cifras asignadas a variables categóricas o de referencia (por ejemplo, el sexo) son solo etiquetas y pueden utilizarse únicamente para distinguir entre grupos. Las cifras adjuntas a escalas ordinales (por ejemplo, para representar las respuestas a un cuestionario que indican el grado de consenso con un enunciado) brindan información sobre valores relativos, aunque en los análisis estadísticos a menudo se las trata como si tuvieran las propiedades de las escalas de intervalo (por ejemplo, la temperatura) o de proporción (por ejemplo, la cantidad de años de experiencia docente).
- Cómo se sumarán los datos de una cantidad de variables para producir una nueva variable y cómo se utilizará esta nueva variable. Por ejemplo, un índice de pobreza podría construirse a partir de variables tales como el ingreso doméstico, la ubicación de la vivienda, la cantidad de habitaciones en la vivienda, la propiedad de la vivienda, la cantidad de hijos y el nivel de educación de los padres. En el diseño del plan se debe considerar cómo se sumarán las variables para representar la pobreza.

REDACCIÓN DE ÍTEMS PARA CUESTIONARIOS

El modo en que está configurado un ítem debe dejar en claro qué información se requiere. Más aún, los encuestados deben poder proporcionar la información. Por lo tanto, no se les pedirá a los alumnos de menor edad que recuerden cuántos días han estado ausentes durante el año escolar, ya que lo más probable es que no puedan recordarlo de manera confiable. A lo sumo, podrían recordar cuántos días no asistieron durante la semana anterior.

La redacción de los ítems debe ser tan simple y clara como sea posible. El vocabulario debe ser familiar y las oraciones breves y directas. A menos que el cuestionario se administre de forma oral, todos los encuestados deben poder leerlo. También es importante que todos los encuestados interpreten los ítems del mismo modo. De lo contrario, resultará difícil la interpretación de los resultados de manera coherente.

La primera parte de un ítem en un cuestionario puede ser una pregunta, una oración incompleta o un enunciado que los encuestados deben evaluar.

El estilo en el que se dirige al encuestado debe ser coherente.

Puede utilizarse cualquiera de los siguientes:

- Referirse a “tú”; por ejemplo, “¿Cuántos años tienes?”
- Referirse a “yo”; por ejemplo, “Yo vengo a la escuela en...”

PREGUNTAS

Las preguntas deben ser claras e inequívocas. La siguiente pregunta es ambigua.

¿Cuánto tiempo hace que es docente?

Esta pregunta confunde el tiempo que ha transcurrido desde la formación con el tiempo dedicado a la enseñanza. Cualquiera que haya dejado la enseñanza y, luego, la haya retomado, por ejemplo, una mujer que se tomó un tiempo para criar a su familia, no estará segura de cómo responder. No está claro si esta pregunta es una medida de la experiencia docente o del tiempo transcurrido desde la formación. Aquí hay al menos dos preguntas:

¿Cuándo finalizó su formación docente?

¿Cuántos años de experiencia docente tiene?

La segunda pregunta aún resulta problemática, ya que no queda claro cómo deberían medirse los años de experiencia docente. Por ejemplo, ¿cinco años de experiencia de trabajo a tiempo parcial, un día a la semana, deberían contarse como cinco años o como el equivalente a un año? Si casi todos los cargos docentes en el país son de tiempo completo, entonces la ambigüedad es improbable, pero si muchos cargos son de tiempo parcial, es probable. La pregunta debería ser la siguiente:

¿Cuántos años de experiencia docente de tiempo completo (o equivalente a tiempo completo) tiene?

Es posible que sea necesario definir lo que se quiere decir con “equivalente”.

ENUNCIADOS

Los ítems que generalmente comienzan con un enunciado requieren que los encuestados realicen alguna evaluación de él, tal como “muy de acuerdo”, “de acuerdo”, “en desacuerdo” o “muy en desacuerdo”.

Los enunciados negativos, tales como “no me gusta la escuela”, deben evitarse, ya que pueden resultar confusos. Si se pregunta a los estudiantes si están de acuerdo o en desacuerdo, aquellos a quienes les guste la escuela deben seleccionar la categoría “en desacuerdo”. Los niños pequeños con frecuencia encuentran difíciles los dobles negativos.

Mantenga los enunciados tan neutros como sea posible. Un enunciado que diga “me gusta la escuela” es mejor que uno que diga “me encanta la escuela”. Los estudiantes pueden expresar mayor afición a la escuela si seleccionan “muy de acuerdo” para su respuesta.

Los enunciados deben centrarse en un tema. Así, deben evitarse los enunciados tales como “me esfuerzo y me va bien en mis tareas”. Los estudiantes a los que les va bien en la escuela sin esforzarse no sabrán qué respuesta elegir. Los estudiantes que se esfuercen mucho posiblemente estén de acuerdo con este enunciado; sin embargo, puede que no les vaya bien en sus tareas. El enunciado se expresa mejor como dos enunciados: “Me esfuerzo en la escuela”; “me va bien en mis tareas”.

CATEGORÍAS DE RESPUESTA

Las buenas categorías de respuesta tienen el mismo significado para todos los encuestados. Las siguientes categorías de respuesta pueden tener distinto significado para diferentes personas:

¿Cuántos libros hay en la biblioteca del aula?

- A. *ninguno*
- B. *unos pocos*
- C. *algunos*
- D. *muchos*

Las categorías de respuesta para el ítem deben cuantificarse de modo que el significado quede claro:

¿Cuántos libros hay en la biblioteca del aula?

- A. *no hay biblioteca*
- B. *de 1 a 10*
- C. *de 11 a 20*
- D. *más de 20*

En ocasiones, las categorías de respuesta pueden tener distintos significados para diferentes encuestados, pero esta diferencia es parte de la información buscada, como lo muestra el siguiente ítem:

¿Qué tan buena es la biblioteca de tu escuela?

A. no hay biblioteca

B. mala

C. adecuada

D. buena

E. excelente

Si el ítem es sobre el nivel de satisfacción del encuestado con respecto a la biblioteca escolar, sin importar ninguna medición objetiva de su calidad, entonces este es un buen ítem. Si el ítem se combina con ítems que cuantifican, por ejemplo, aproximadamente cuántos estantes de libros o instalaciones de computadoras tiene la biblioteca, entonces la percepción del encuestado puede compararse con mediciones más objetivas de las instalaciones de la biblioteca.

Las categorías de respuesta deben tener en cuenta el nivel de precisión de respuesta que es posible que puedan dar los encuestados. Es probable que los encuestados desconozcan la cantidad exacta de libros en una biblioteca, a menos que esta sea muy pequeña.

Las categorías de respuesta deben cubrir todas las respuestas posibles. Si existen categorías principales y otras menores, es preferible enumerar las principales e incluir la opción "otros". La realización de una prueba previa ayuda a identificar las categorías principales.

Las categorías de respuesta no deben superponerse ni dejar lagunas. Ambos errores se muestran en la siguiente pregunta:

¿Cuánto tiempo hace que es docente en esta escuela?

A. menos de 5 años

B. menos de 10 años

C. más de 10 años

Los profesores con menos de 5 años de experiencia docente no saben si deben elegir la primera opción o la segunda, mientras que aquellos con 10 años de experiencia docente no tienen una opción que elegir.

Los ítems de los cuestionarios que comienzan con un enunciado que los encuestados deben evaluar pueden tener muchas categorías de respuesta diferentes. Es importante que estas no se superpongan. Algunos ejemplos de categorías de respuesta usadas en el cuestionario para docentes del Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) son los siguientes:

- Sí, no.
- Muy de acuerdo, de acuerdo, en desacuerdo, muy en desacuerdo.
- Casi todos los días, una o dos veces a la semana, una o dos veces al mes, nunca o casi nunca.
- En absoluto, un poco, bastante, mucho.
- No es importante, un poco importante, muy importante.

TRATAMIENTO DE TEMAS DELICADOS

Algunos temas son delicados, p. ej., si los docentes tienen un segundo empleo para complementar sus salarios. Si es probable que la mayoría de los encuestados no responda con honestidad a una pregunta, debería excluirse. Es posible que los responsables políticos estén muy interesados en esta información, pero no tiene mucho sentido recopilar datos que no son confiables. A veces, en su lugar se puede recopilar información que no sea tan delicada.

Es posible que al incluir preguntas sobre temas delicados, los encuestados se sientan ofendidos y no deseen responder el resto de los ítems o que devuelvan el cuestionario. Si hay dudas sobre la sensibilidad de los temas, lo mejor es no incluirlos.

DISEÑO DEL CUESTIONARIO

Hay dos consideraciones importantes en cuanto al diseño y formato de los cuestionarios: (a) facilidad de uso para el encuestado y (b) facilidad de uso para el procesamiento de datos. Los cuestionarios son fáciles de utilizar cuando cumplen con las siguientes características:

- Una manera simple y coherente de responder a las preguntas.
- Una presentación ordenada.

- Preguntas separadas fácilmente identificables.
- Categorías de respuesta que se asocian claramente con cada pregunta.
- Encabezados, fuentes y formato coherentes.
- Categorías de respuesta codificadas para el ingreso de los datos.

Las categorías de respuesta pueden establecerse de diferentes maneras. Pueden figurar en una columna, de manera vertical o en una fila horizontal. Los encuestados pueden marcar una letra o un número con un círculo, o pueden tildar una casilla para indicar su elección. Es preferible tener un estilo de respuesta coherente.

En el recuadro 9.1 se brinda un ejemplo de un ítem en el que las respuestas no se identifican claramente con las categorías de respuesta. El recuadro 9.2 muestra un ajuste mejor.

REVISIÓN DE LOS CUESTIONARIOS

Redactar cuestionarios es mucho más difícil de lo que parece. Todos los ítems deben ser cuidadosamente examinados y revisados para garantizar que sean claros e inequívocos. Es muy recomendable consultar a un comité para revisar el cuestionario. Entre los miembros del comité debe haber redactores de ítems, personas familiarizadas

RECUADRO 9.1

Correspondencia deficiente entre las casillas y las categorías de respuesta

El siguiente ejemplo muestra una correspondencia deficiente entre las casillas y las categorías de respuesta:

¿Cuánto tiempo te toma todos los días llegar a la escuela?

menos de 15 minutos 15 minutos 30 minutos
45 minutos 1 hora más de 1 hora

Las casillas se encuentran entre las categorías de respuesta en lugar de estar claramente alineadas con cada categoría.

RECUADRO 9.2**Mejor correspondencia de casillas y categorías de respuesta**

El siguiente ejemplo muestra una buena correspondencia entre las casillas y las categorías de respuesta:

¿Con qué frecuencia las siguientes personas te ayudan con la tarea de la escuela para el hogar?

	<i>Nunca o casi nunca</i>	<i>Unas pocas veces al año</i>	<i>Cerca de una vez por mes</i>	<i>Varias veces por mes</i>	<i>Varias veces por semana</i>
<i>a) Tu madre. . .</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>b) Tu padre. . .</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>c) Tus hermanos y hermanas. . .</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>d) Uno de tus abuelos. . .</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

con las características de la población encuestada y personas que sean capaces de garantizar que los ítems sean culturalmente adecuados. Actualmente, algunas evaluaciones nacionales e internacionales controlan la interpretación que los encuestados hacen de los ítems antes de su administración principal.

Es de utilidad que los miembros del comité intenten completar el cuestionario como si fueran los encuestados. Este proceso ayudará a identificar si hay categorías de respuesta que podrían no ser claras, que podrían superponerse o que podrían no incluir algunas clases de respuesta.

Los miembros del comité deben criticar los ítems, en especial, en lo referido a su claridad de expresión y la pertinencia de las categorías de respuesta. Deben garantizar que la expresión sea lo más clara y simple posible, que el estilo de los ítems sea coherente y que se los presente en un orden lógico con instrucciones adecuadas.

Los miembros del comité deben verificar que los ítems se correspondan con el proyecto de cuestionario y garantizar que cada ítem brinde la información requerida. También es necesario que se verifique

que la cantidad de ítems utilizados sea adecuada para medir cada variable con suficiente precisión.

Una vez mejorado el cuestionario, se debe dar a los responsables políticos la oportunidad de revisarlo. Es necesario que los responsables políticos aprueben los ítems, en especial si tratan temas políticamente delicados. Los responsables políticos también deben verificar que los ítems aporten información de utilidad.

Después de la mejora del cuestionario, debe realizarse una prueba previa o una prueba de campo junto con los materiales del examen. La prueba previa brinda una oportunidad de mejorar la calidad de los ítems, y reduce el tiempo y los costos del procesamiento de datos del cuestionario final. Los ítems que no funcionan (por ejemplo, aquellos en los que se confunden los encuestados) pueden quitarse y las categorías de respuesta ampliarse o resumirse.

Luego de la administración del cuestionario, el examinador debe recabar observaciones de los encuestados (estudiantes o docentes) sobre los ítems que no son claros o que no contienen la información adecuada. Los examinadores deben verificar que ningún ítem resulte ofensivo por tocar temas delicados.

Un análisis estadístico más formal de las respuestas puede indicar que las categorías de respuesta deben diferenciarse más. Por ejemplo, si la mayoría de los estudiantes selecciona una categoría de respuesta en particular para un ítem. La categoría debe separarse en categorías más diferenciadas para obtener información más precisa.

Si el rango de respuestas posibles para un ítem es potencialmente muy amplio y difícil de prever, el ítem debe dejarse como de respuesta abierta en la prueba previa. Posteriormente, las respuestas pueden clasificarse y utilizarse para generar categorías para un ítem de elección forzada en el cuestionario final.

El CD complementario contiene ejemplos tomados de cuestionarios para estudiantes, docentes, directores y padres de evaluaciones nacionales e internacionales.

También incluye ejemplos de guías de corrección para respuestas abiertas.



CODIFICACIÓN DE LAS RESPUESTAS DEL CUESTIONARIO

Las categorías de respuesta deben codificarse para la captura de datos. La codificación puede ser alfabética o numérica.

Los códigos alfabéticos generalmente requieren que los encuestados marquen una letra con un círculo para su respuesta. Este método puede no ser apropiado para los alumnos de menor edad. El tildar casillas o sombrear círculos puede resultar más fácil para las personas con capacidades de comprensión lectora limitadas. Si los ítems utilizan este sistema, deben codificarse de manera numérica.

Si se utiliza la codificación numérica, la primera categoría de respuesta se codifica como 1, la segunda categoría como 2 y así sucesivamente. La captura de datos resulta más eficaz si los códigos están impresos en los cuestionarios. Puede utilizarse una fuente pequeña en escala de grises, como se muestra en el recuadro 10.1. En el ejemplo, las categorías de respuesta se numeran debajo de las casillas: caminando es la categoría 1, en transporte público es la categoría 2 y así sucesivamente. Los estudiantes marcan las casillas que les parecen adecuadas. La persona que ingresa los datos coloca el número de la casilla que el estudiante seleccionó.

Si los encuestados tienen la oportunidad de seleccionar más de una categoría de respuesta para un ítem, cada categoría se debe considerar

RECUADRO 10.1**Ejemplo de codificación en escala de grises**

Hoy vine a la escuela

caminando en transporte público en transporte privado montando un animal.

1

2

3

4

RECUADRO 10.2**Ejemplo de tratamiento de los ítems como categorías individuales para la captura de datos**

Si estuvo ausente la semana pasada, marque una o más casillas para mostrar el motivo.

Estuve enfermo.

1

Tuve que ayudar a mis padres.

1

Había mal clima.

1

No tenía alimentos.

1

Mi familia tuvo problemas.

1

No tenía un uniforme limpio o ropa apropiada.

1

No era seguro (problemas relacionados con la paz y el orden).

1

Otros. _____

1

como un ítem individual, tanto para la captura de datos como para su procesamiento. Este procedimiento posibilita un seguimiento de las categorías que seleccionó cada encuestado. El ítem del recuadro 10.2 se presenta al encuestado como una pregunta con múltiples respuestas posibles; sin embargo, a los efectos de la captura de datos, se considera

como ocho ítems individuales. Las respuestas a la primera categoría (ausente por enfermedad) se registran como 1 o ausente; las respuestas a la segunda categoría (ayudó a sus padres) se registran como 1 o ausente; las respuestas a la tercera categoría (mal clima) se registran de manera similar y así sucesivamente para cada una de las ocho categorías.

PREPARACIÓN DE LOS CUESTIONARIOS PARA LA CAPTURA DE DATOS

Los datos del cuestionario se pueden escanear con un equipo especial o pueden ingresarse manualmente. El diseño y la disposición del cuestionario tendrán que personalizarse si se va a utilizar un escáner.

Si la captura de datos se realiza manualmente, se puede ingresar la información directamente desde el cuestionario si se han codificado las categorías de respuesta. Sin embargo, las personas que realizan la captura de datos pueden tener dificultades para mantener un nivel alto de exactitud, especialmente si no están familiarizados con este tipo de trabajo. También es probable que la exactitud se vea afectada si la disposición de los ítems varía en gran parte o si algunos ítems tienen una gran cantidad de categorías de respuestas.

La captura de datos será más fácil si los evaluadores y correctores escriben el código de la categoría seleccionada sobre el margen izquierdo cerca de cada número de ítem. De esta manera, la captura de datos se convierte simplemente en el ingreso de los códigos escritos en el margen. Agregar casillas ligeramente sombreadas en el margen para que evaluadores y correctores escriban los códigos hará que el proceso resulte más eficiente.

RESPUESTAS AMBIGUAS O FALTA DE CÓDIGO

En ocasiones, los encuestados no responden los ítems o los responden de manera ambigua, por ejemplo, si seleccionan más de una categoría de respuesta cuando las categorías son mutuamente excluyentes.

La recopilación de información sobre la falta de respuestas indica si los encuestados fallaron regularmente en responder a algunos ítems. Por ejemplo, es posible que el cuestionario sea demasiado extenso, de modo que los ítems que se encuentran al final no se hayan respondido, o es posible que un ítem esté muy cerca de otros y pueda pasarse por alto con facilidad. La recopilación de información sobre respuestas ambiguas también podrá indicar si un ítem posiblemente no ha resultado claro para muchos encuestados o si estos no comprendieron cómo debían completar el cuestionario.

La persona que ingresa los datos debe saber cómo codificar la falta de respuestas o las respuestas ambiguas. Los códigos para la falta de respuesta o para respuestas ambiguas no deben confundirse con los códigos que se utilizan para las categorías de respuesta.

Para indicar falta de respuesta o ningún intento puede utilizarse una letra del alfabeto, como la X. El código para respuestas ambiguas puede ser una segunda letra, como la Y. Los ítems de opción múltiple en los exámenes utilizan los códigos 9 para “ausente” y 8 para la selección de dos opciones o más. Estos códigos generalmente no se utilizan para los cuestionarios ya que algunos ítems de los cuestionarios probablemente tendrán 8 o 9 categorías de respuesta.



COTEJO DE LOS CUESTIONARIOS CON LOS DATOS DE LA PRUEBA

La manera en la que se hacen coincidir los datos del cuestionario y de la prueba se orientará según las necesidades del análisis y el informe. Todas las coincidencias deben establecerse de manera clara e inequívoca antes de la recopilación de los datos. Todo error de coincidencia que se descubra después de haber recopilado los datos puede ser difícil de corregir o incluso imposible. Estos errores pueden hacer que haya que abandonar algunos análisis planificados.

CUESTIONARIOS PARA ESTUDIANTES

La manera más fácil de que los datos de los cuestionarios y las pruebas de los alumnos coincidan es imprimir las pruebas y los cuestionarios en un solo cuadernillo. Los alumnos registran sus nombres en el cuadernillo y el examinador se asegura de que el alumno trabaje en su propio cuadernillo en cada una de las sesiones de la prueba.

Si las pruebas y los cuestionarios son documentos separados, uno de los métodos para que los datos coincidan es sobreimprimir o etiquetar tanto las pruebas como los cuestionarios con el nombre de

los estudiantes. Los nombres se toman de la lista de inscritos de la escuela y su redacción debe ser idéntica en cada etiqueta. Nuevamente, el examinador debe asegurarse de que los estudiantes trabajen en las pruebas y cuestionarios que presentan sus nombres claramente etiquetados.

Si no es posible etiquetar con anterioridad los cuadernillos y cuestionarios individuales, los cuestionarios de los estudiantes deben contener la información identificadora suficiente que permita relacionarlos con los datos del examen. Es preferible asignar una identidad (ID) numérica a los estudiantes y asegurarse de que utilicen el mismo número identificativo en cada cuadernillo y cuestionario. El examinador deberá supervisar este procedimiento. El analista de datos también requerirá la lista de los nombres de los estudiantes y de los números identificativos, ya que se podrá recurrir a la lista de nombres como respaldo en caso de que ocurran errores de identificador.

No es conveniente depender de los nombres de los estudiantes para identificar los formularios. A menos que los nombres sean completamente idénticos en cada formulario y que la persona que procesa los datos los ingrese de manera idéntica, sin errores de escritura, la computadora no podrá reconocerlos. La identificación tendrá que realizarse, en ese caso, manualmente, una tarea cara y que lleva mucho tiempo. Algunos estudiantes harán que resulte más complicada la identificación por nombres al usar nombres diferentes (tales como formas abreviadas, apodos o nombres religiosos) en distintos formularios, escribir de manera ilegible en uno o más formularios, o no escribir sus nombres en uno o más formularios.

CUESTIONARIOS PARA PADRES

Los datos de los cuestionarios para padres generalmente se correlacionan con los datos de los estudiantes. Es probable que la correlación se realice mediante el nombre de los estudiantes. Se presenta el mismo problema que se describió para el cuestionario de los estudiantes. Se deben establecer procedimientos que aseguren la coherencia.

CUESTIONARIOS PARA DOCENTES Y DIRECTORES

Los cuestionarios para docentes y directores generalmente se correlacionan solo con el grado y la escuela. Si se conoce el grado del estudiante, puede utilizarse la información del docente en el análisis de los datos del estudiante. El examinador debe controlar que los docentes y directores hayan brindado esta información en los cuestionarios.

Al recibirse los cuestionarios cumplimentados por las escuelas, los correspondientes a cada escuela deben guardarse en un paquete aparte. De este modo, incluso si no se proporcionó la información de la escuela en algunos cuestionarios, se puede obtener la información pertinente de otros cuestionarios del mismo paquete.

PARTE

3

DISEÑO DE UN
MANUAL PARA LA
ADMINISTRACIÓN
DE LA PRUEBA

MANUAL DEL
EXAMINADOR

Es necesario un manual para guiar la administración de la prueba, que debe ser estandarizado para que todos los estudiantes realicen la prueba en las mismas condiciones. El propósito principal del manual es especificar las condiciones exactas según las cuales se debe llevar a cabo la prueba, entre ellas los requerimientos de preparación y los procedimientos para garantizar la seguridad. Los estudiantes que realicen la prueba deben responder a las mismas preguntas de práctica y recibir las mismas indicaciones sobre cómo mostrar sus respuestas. A todos se les debe dar la misma cantidad de tiempo para realizar la prueba con el mismo grado de supervisión.

El desempeño de los estudiantes en una evaluación nacional debe ser una medida de su capacidad para responder los ítems sin apoyo externo. Los estudiantes deben comprender lo que deben hacer y cómo mostrar sus respuestas, pero no se les debe dar ninguna otra ayuda ni pueden tener acceso a otros recursos que no formen parte de la prueba. Seguir con los procedimientos establecidos en el manual de administración contribuirá a asegurar que así sea. El CD complementario contiene ejemplos de administración de pruebas y de manuales de coordinación escolar.

CONTENIDOS DEL MANUAL

Los manuales de administración deben brindar información que responda a cada una de las siguientes preguntas:

- *¿Para qué se realiza la prueba?*
 - Una breve explicación del propósito de la prueba y la forma en la que se utilizarán los datos.
- *¿Qué pruebas se realizarán, qué estudiantes se evaluarán y cuándo serán evaluados?*
 - Qué pruebas se administrarán en la escuela.
 - Qué estudiantes realizarán cada prueba.
 - Fecha y hora de la administración de la prueba.
 - Orden de administración de las pruebas.
 - Duración de la administración de cada prueba.
 - Toda pausa necesaria entre las administraciones de las pruebas.
 - Toda posibilidad de flexibilidad en el cronograma de administración.
- *¿Qué materiales se necesitan para la prueba?*
 - Lista de todos los materiales que se proporcionan para la prueba.
 - Cantidades de cada uno de los materiales proporcionados para la prueba, tales como uno por estudiante o uno por docente.
 - Lista de los materiales que debe proporcionar la escuela, tales como lápices y borradores.
- *¿Cómo debe disponerse la sala para la prueba?*
 - Instalaciones materiales que debe proveer la escuela, tales como escritorios y sillas.
 - Deben quitarse de la sala o cubrirse los recursos que podrían ayudar a los estudiantes, tales como cuadros de las tablas de multiplicar o carteles que muestren reglas gramaticales.
- *¿Qué preparación se necesita?*
 - Cómo podría el director motivar al personal y a los estudiantes a apoyar la administración de la prueba antes de que se lleve a cabo.
 - La información que podría necesitar el examinador, tal como una lista de los nombres de la clase.

- Cómo deberían clasificarse, numerarse o nombrarse los cuadernillos de la prueba para que estén listos para utilizarse.
- Cómo deberían organizarse los grupos de alumnos para la prueba.
- *¿Cómo se debe llevar a cabo la prueba?*
 - La manera en la que los alumnos deben escribir los nombres en los cuadernillos de la prueba y registrar la información general en la portada.
 - Cuándo y cómo el administrador debe controlar que los alumnos hayan registrado la información en la portada del cuadernillo de la prueba.
 - Cómo deben administrarse y explicarse las preguntas de práctica.
 - Qué indicaciones sobre la prueba deben recibir los estudiantes.
 - Qué nivel de apoyo puede ofrecer el examinador durante la prueba.
 - Cuánto tiempo tienen los estudiantes para completar la prueba.
 - Qué condiciones debe mantener el examinador durante la prueba.
 - Quién podrá ingresar a la sala durante la administración de la prueba.
- *¿Cómo se guardarán los materiales de la prueba?*
 - Procedimientos para garantizar la seguridad del material de la prueba antes, durante y después de la misma.
- *¿A quién se puede contactar para solicitar ayuda?*
 - Datos de contacto de personas que pueden brindar ayuda con problemas o proporcionar información adicional.

Es posible que se incluya otra información en el manual de administración para agilizar el traslado de los cuadernillos de prueba dentro y fuera de las escuelas. Es probable que esta información varíe según si las pruebas son administradas por una agencia externa o por los docentes de la escuela.

USO DEL MANUAL

Tanto el director de la escuela como el examinador deben utilizar el manual. Algunas evaluaciones preparan manuales diferentes para

directores o para quienes tienen la responsabilidad general de llevar a cabo las actividades de la evaluación nacional dentro de cada escuela.

El director necesita el manual para garantizar que su escuela esté adecuadamente preparada para la administración de la prueba. Debe saber lo suficiente sobre la prueba como para alentar al personal y a los estudiantes a apoyar su realización, y para alentar a los estudiantes a que se esfuercen tanto como puedan. El director o el coordinador de la escuela para la evaluación nacional (si se ha designado uno) debe contar con información suficiente como para poder organizar la escuela y asegurarse de que los alumnos indicados estén disponibles en el momento requerido y con el material adecuado, que tengan espacio suficiente para realizar la prueba y que los materiales de la prueba puedan guardarse de manera segura.

Los examinadores necesitan el manual para indicarles exactamente qué es lo que deben hacer para administrar la prueba de manera correcta, y cuándo y cómo administrarla. Deben controlar que haya suficiente disponibilidad de materiales para la prueba y que se seleccione a los estudiantes adecuados para realizarla. Deben saber qué información sobre la prueba brindarle a los estudiantes, cómo explicarles las preguntas de práctica y de cuánto tiempo disponen los estudiantes para realizar la prueba. También deben saber qué procedimientos de seguridad seguir para guardar los materiales de la prueba.

CARACTERÍSTICAS DEL MANUAL

Un buen manual contiene toda la información necesaria y es fácil de utilizar. La información está ordenada de manera lógica, las instrucciones son claras y están completas, y el lenguaje es simple y directo. La organización de la información en viñetas, casillas o tablas hará que sea más fácil de leer. Un buen manual debe tener un índice de contenidos con títulos claros (consulte el recuadro 12.1).

RECUADRO 12.1**Instrucciones del manual de administración**

En una evaluación nacional, aparecía la siguiente información en una fuente grande (Arial 14) y ocupaba la portada completa del manual de administración:

Lea este manual de administración antes de que los estudiantes realicen la prueba.

Los estudiantes deben realizar esta prueba en DOS DÍAS.

- La prueba se divide en cuatro sesiones. Los estudiantes deben realizar dos sesiones cada día.
- Los estudiantes deben tener una pausa entre cada sesión.
- No permita que los estudiantes lleven a cabo toda la prueba sin interrupción.

Reglas de administración

- Los docentes deben supervisar todas las sesiones en todo momento.
- NO está permitido que los estudiantes se lleven el cuadernillo de la prueba fuera de la sala o que trabajen en él después de que se haya marchado el docente.
- Los estudiantes deben utilizar los lápices proporcionados con borrador en el extremo.
- Al realizar la prueba, no deben utilizar materiales del aula, tales como cuadernillos de trabajo, diccionarios o calculadoras.
- No deben recibir ayuda para responder las preguntas. Por ejemplo, si un estudiante no entiende qué debe hacer, explique las preguntas de práctica otra vez e indíquele que haga su mejor intento, pero no brinde otro tipo de ayuda.

Seguridad de la prueba

- Los materiales de la prueba deben **GUARDARSE DE MANERA SEGURA EN TODO MOMENTO**.
- Los cuadernillos de la prueba de los estudiantes **NO** deben copiarse por ningún motivo.
- Los estudiantes **NO** deben llevarse a casa los cuadernillos de la prueba.

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

¿CUÁNTA INFORMACIÓN SE NECESITA?

La información sobre las condiciones generales de administración de la prueba y la preparación de los materiales debe ser al mismo tiempo exhaustiva y tan breve como sea posible (consulte el recuadro 12.2).

Las instrucciones que el examinador da a los estudiantes deben ser totalmente por escrito. Todo lo que el examinador deba decirle a los estudiantes sobre la prueba, las preguntas de práctica o las condiciones de administración de la prueba debe estar en forma escrita. El examinador debe leer del manual sin modificar las frases. Este procedimiento garantiza que todos los estudiantes que realizan la prueba reciban exactamente las mismas instrucciones.

RECUADRO 12.2

Información para docentes y directores

La información sobre los materiales de la prueba debe ser concisa y debe describirse de modo que resulte fácil de verificar. El siguiente extracto de una evaluación a gran escala que se realizó en Papúa Nueva Guinea indica al director qué materiales se han enviado a la escuela y cómo saber qué clases participarán en la prueba:

Materiales de la prueba

Su inspector superior de escuela primaria le dirá qué clases de su escuela deben participar en esta prueba.

Debe haber recibido los siguientes materiales:

- una carta de presentación para el director
- un cuadernillo de prueba para cada estudiante que participa
- un manual de administración para cada docente que administra la prueba
- un cuestionario de antecedentes docentes para cada docente que administra la prueba
- un lápiz con borrador en el extremo para cada estudiante que participa

Si falta algún material o no cuenta con el material suficiente, comuníquese con el inspector superior de escuela primaria.

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

PREGUNTAS DE PRÁCTICA

El recuadro 12.3 muestra instrucciones de procedimiento generales y una descripción del propósito de las preguntas de práctica. Las instrucciones que da el administrador aparecen escritas y resaltadas en un cuadro sombreado. El administrador debe leer estas instrucciones tal como están impresas. La ilustración del dinero que forma parte de las preguntas de práctica en el cuadernillo también aparece en el manual de administración para que el examinador sepa qué ven los estudiantes sin tener que leer del manual y de uno de los cuadernillos de los estudiantes al mismo tiempo.

RECUADRO 12.3

Administración de los ítems de práctica

El siguiente extracto muestra parte de las instrucciones para el planteamiento de algunas preguntas de práctica:

Día uno: Sesión 1

PREGUNTAS DE PRÁCTICA DE MATEMÁTICAS PARA 3.º GRADO
(aproximadamente 10 minutos)

Asegúrese de que cada estudiante tenga el cuadernillo de la prueba con su nombre escrito en la portada. Las preguntas de práctica son para demostrar a los estudiantes las distintas maneras en las que mostrarán sus respuestas.

Pida a los estudiantes que abran sus cuadernillos en Preguntas de práctica de matemáticas (primera hoja).

Diga

Hoy vamos a trabajar en matemáticas para poder averiguar todo lo que ustedes pueden hacer en matemáticas. Primero, resolveremos unas preguntas de práctica para que sepan qué hacer y cómo mostrar sus respuestas.

Sostenga el cuadernillo de un estudiante y señale las preguntas de práctica. Verifique que todos hayan encontrado las preguntas indicadas.

(continúa)

Diga

Veamos la pregunta de práctica uno. Se las leeré.

Aquí hay una imagen de dinero.



¿Cuánto dinero hay en total?

¿Son 2 toeas, 7 toeas, 25 toeas o 205 toeas?

Coloreen el círculo pequeño al lado de la respuesta correcta.

Coloreen solo un círculo.

Espera hasta que todos los alumnos hayan terminado y, luego, verifique sus respuestas.

Diga

La respuesta es 25 toeas. Debían colorear el círculo pequeño al lado de 25 toeas. Si cometieron un error, bórrenlo y coloreen en la respuesta correcta.

De ser necesario, espere hasta que todos los alumnos hayan corregido su trabajo.

Fuente: Departamento de Educación de Papúa Nueva Guinea 2004.

ENSAYO

El manual debe estar preparado para ensayarse en la prueba previa o de campo de los ítems de la prueba. Al realizar una prueba previa, se destacarán las ambigüedades o los malentendidos que requieran un

esclarecimiento o una mejora en la versión final. Debido a que las condiciones de la prueba previa o de campo deben ser lo más parecidas posible a las de la administración final, el manual debe estar tan terminado como sea posible al momento de esa prueba.

Las instrucciones generales sobre la administración de la prueba generalmente pueden redactarse en cualquier momento después de finalizado el proyecto. Los proyectos deben especificar todos los requisitos relacionados con la cantidad de pruebas y su duración, y qué alumnos deben realizarlos. Durante la prueba previa, el administrador debe recopilar información como la siguiente para ayudar al encargado del desarrollo de la prueba a mejorar la prueba final:

- Si los estudiantes necesitaron todas las preguntas de práctica, si estas fueron suficientes y si las explicaciones fueron suficientemente claras.
- Si la prueba fue de la extensión adecuada o si fue demasiado larga, y aproximadamente cuántos estudiantes la terminaron 10 minutos antes (si en la misma clase se utilizan distintos formularios, el examinador puede comparar el tiempo que los estudiantes necesitaron para cada formulario).
- Si los estudiantes parecieron concentrados en la prueba.
- Si los estudiantes tuvieron recursos adecuados y suficientes, tales como lápices o borradores.
- Si las instalaciones de la escuela fueron adecuadas para llevar a cabo la prueba.

REVISIÓN

El encargado del desarrollo de la prueba y los redactores de los ítems son los responsables de la revisión y la mejora de las instrucciones de las preguntas de práctica. Las preguntas de práctica y las instrucciones de administración deben darse a la misma clase de comité que se establece para revisar los ítems.

El encargado del desarrollo de la prueba y la persona responsable de la logística de su distribución y producción deben revisar y mejorar los procedimientos para el traslado de los cuadernillos de prueba dentro y fuera de las escuelas.

Varias personas con formación similar a la de los examinadores también deben revisar todo el manual, con el fin de verificar que las instrucciones sean claras y esclarecer toda ambigüedad que pudiera surgir.

Como todos los materiales de prueba, el manual se debe corregir con regularidad y de manera exhaustiva por expertos en estas tareas. Las instrucciones para las preguntas de práctica y para la administración de la prueba solo se pueden corregir adecuadamente si los correctores cuentan también con copias de los cuadernillos de la prueba de los estudiantes correspondientes.



ELECCIÓN DEL EXAMINADOR

Las personas deben estar seguras de que la prueba se administró bajo condiciones estandarizadas. Los examinadores deben ser considerados confiables.

La elección del examinador depende de las condiciones existentes en cada país. Los inspectores escolares pueden ser examinadores ideales en algunos países y problemáticos en otros. Si los inspectores ven la administración de la prueba como una tarea adicional fuera de la descripción de su trabajo, que consume recursos escasos o que es de poco interés para ellos, es posible que no estén motivados para realizar correctamente la tarea.

En algunas evaluaciones nacionales se utilizan examinadores externos. En el mejor de los casos, se trata de personas que pueden seguir instrucciones con exactitud, que tienen el tiempo y los recursos para realizar la tarea adecuadamente y que no tienen intereses particulares en el resultado de la prueba, más que su correcta administración. En Timor Oriental, por ejemplo, se capacitó y se pagó a censistas para que administraran una evaluación nacional en las escuelas. Eran personas que comprendían la importancia de recopilar los datos de manera sistemática y no tenían ningún interés en el desempeño de

los estudiantes. Su trabajo se supervisó para garantizar que fuese de un estándar adecuado.

En algunos países, la administración de una evaluación nacional por parte de los docentes comprometería gravemente la credibilidad de los datos, pero en otros es posible que no. El problema principal de pedir a los docentes que administren la prueba es que pueden ofrecer ayuda a los estudiantes, intencionalmente o no. Hay muchas razones posibles para este fenómeno. A algunos docentes puede preocuparles que los datos de la prueba se utilicen para juzgar su desempeño como docentes. Pueden sentir que deben ayudar a los estudiantes para, así, mantener su trabajo o su estatus profesional. Los directores pueden sentir que su estatus está incluso más amenazado. Algunos docentes pueden creer que la prueba no es una medición justa de su enseñanza o del desempeño de los estudiantes y pueden sentirse obligados a brindar ayuda para hacer que la prueba sea “justa”. Algunos docentes pueden tener las mejores intenciones de administrar la prueba como se indica, pero es posible que no puedan salirse de su rol de docente. Es posible que ayuden a los estudiantes sin darse cuenta de lo que hacen o porque no soportan ver que tienen dificultades sin ofrecerles ayuda.

El volumen 3 de esta serie, *Implementación de una evaluación nacional del rendimiento académico*, aborda la elección de examinadores y describe algunas de las ventajas y desventajas de los diferentes tipos de examinadores.

CUMPLIMIENTO DE LAS INSTRUCCIONES

El manual debe distinguir entre instrucciones específicas que deben seguirse al pie de la letra e instrucciones más generales que permiten al examinador cierto grado de adaptación a las condiciones de la clase. El examinador no debe desviarse de las instrucciones específicas. Un ensayo previo del manual ayudará a identificar cualquier error o ambigüedad en las instrucciones.

Los examinadores deben ayudar a los alumnos solo a entender lo que deben hacer y cómo expresar sus respuestas. Deben dejar en claro que no pueden ayudar a ningún alumno a responder las preguntas.

Tampoco deben ofrecer ayuda para interpretar alguna pregunta, explicar el significado de alguna palabra o sugerir modos en los que el estudiante podría intentar responder una pregunta. Si un estudiante solicita ayuda, el examinador debe decirle al estudiante que intente hacer lo mejor que pueda.

Los examinadores no deben traducirles la prueba en otro idioma a los estudiantes, a menos que las instrucciones autoricen específicamente esta función.

En algunas pruebas, los examinadores pueden leerles las preguntas a los alumnos. El examinador debe leer en forma pausada y clara toda la prueba en voz alta a la clase, pregunta por pregunta, o puede leer preguntas aisladas según lo soliciten los estudiantes. En todo caso, el examinador debe leer las palabras exactas de la pregunta en el idioma en que estén impresas.

Los examinadores deben tener un reloj. Deben tomar nota en la pizarra o en un papel de la hora exacta a la que comienza la prueba y la hora a la que finaliza. Deben garantizar que los estudiantes conozcan el tiempo del que disponen para realizar la prueba. Generalmente, esto implica decirle a los estudiantes de cuánto tiempo disponen al comienzo y hacerles un recordatorio cuando falten 10 minutos, 5 minutos o cuando queden 2 minutos, según la extensión de la prueba.

Los examinadores deben alentar en voz baja a los estudiantes a que intenten completar la prueba si resulta obvio que emplean demasiado tiempo en una pregunta y son reticentes a avanzar. Pueden hacerlo susurrando al estudiante que escriba su mejor respuesta y que, luego, intente la siguiente pregunta.

Durante la administración de la prueba, solo se permiten en la sala los materiales que se especifican en el manual. Generalmente, los estudiantes traen al examen sus propios lápices y borradores. No se deben permitir cartucheras ni mochilas. Todo lo que pueda ayudar a los estudiantes a responder las preguntas de la prueba debe sacarse de la sala. Los estudiantes no deben tener acceso a recursos tales como diccionarios o calculadoras, a menos que las condiciones de la prueba específicamente lo autoricen.

El examinador, los estudiantes que participan y, posiblemente, un supervisor deben ser las únicas personas en la sala durante la

administración de la prueba. Ni el director ni ningún otro docente deben poder caminar por la sala. Debe notificarse al encargado de la prueba cualquier cambio inevitable en las condiciones de administración de la prueba.

Durante la administración de la prueba, el examinador deberá recopilar información sobre cualquier variación que ocurriera en las condiciones de administración para estudiantes en particular. Con frecuencia, la portada de los cuadernillos de la prueba tendrá espacio para que el examinador indique que los estudiantes estuvieron ausentes en toda la prueba o en una parte de esta. Si un estudiante tiene que salir de la sala más temprano por enfermedad y no completa la prueba, el examinador debe registrar esta información.

El CD que acompaña este libro contiene ejemplos de instrucciones generales y específicas para los examinadores. También brinda algunas sugerencias sobre cómo diseñar un manual de administración de la prueba.

CONTROL DE CALIDAD

Los examinadores deben seleccionarse por su aptitud para la tarea. Deben hablar el idioma en el que está escrito el manual con fluidez. Además, deben comprometerse a realizar bien su tarea.

Independientemente de su nivel de antigüedad en el empleo o sus títulos académicos, los examinadores necesitan una capacitación. Deben asistir a una sesión de capacitación que explique el propósito de la prueba y su función en la administración de esta. Deben entender por qué es importante seguir las instrucciones y se les debe dar la oportunidad de ensayar cómo administrar la prueba con otros colegas examinadores. Deben tener la oportunidad de formular preguntas sobre los procedimientos descritos en el manual.

Si los docentes van a administrar la prueba a sus propios alumnos, la capacitación debe garantizar que entiendan el propósito de la prueba y se les debe tranquilizar asegurándoles que los datos no se utilizarán para juzgarlos. Deben entender la importancia de no ayudar a los alumnos a responder las preguntas.

Los examinadores deben ser supervisados por lo menos en algún momento mientras administran la prueba. Puede que no sea posible supervisar a todos, pero debe ser factible controlar al azar a algunos examinadores.

También puede ser que se les pida a los examinadores que completen y firmen una lista de verificación de sus tareas con el fin de garantizar que han cumplido su trabajo.

LISTA DE VERIFICACIÓN DEL EXAMINADOR

Los detalles de lo que debe figurar en la lista de verificación del examinador podrán variar, según quién administre la prueba y los procedimientos desarrollados para seguir los cuadernillos y garantizar la seguridad. El recuadro 13.1 brinda un ejemplo de una lista de verificación de administración utilizada en Filipinas. Otro ejemplo puede observarse en el volumen 3 de esta serie, *Implementación de una evaluación nacional del rendimiento académico*.

RECUADRO 13.1

Lista de verificación de administración: Un ejemplo de Filipinas

El examinador debe marcar cada ítem para mostrar que lo ha completado y firmar el formulario al final.

Nombre		Fecha	
Tarea	Referencia	Hora	Completado
1. Completar el formulario de asignación de cuadernillos de la prueba de los estudiantes (STBA) mediante la inserción del número de examen en orden consecutivo y de los nombres de los alumnos en orden alfabético.	Formulario STBA	10 min	<input type="checkbox"/>
2. Administrar el cuestionario a los docentes.	Formulario de cuestionario para docentes	15 min	<input type="checkbox"/>

(continúa)

RECUADRO 13.1 (continúa)

Nombre		Fecha	
Tarea	Referencia	Hora	Completado
3. Completar el formulario de observaciones.	Formulario de observaciones del docente	10 min	<input type="checkbox"/>
4. Distribuir los exámenes asignados a cada estudiante y marcar ausente a los estudiantes que no estén presentes.	Formulario STBA	10 min	<input type="checkbox"/>
5. Leer la introducción de <i>Lineamientos</i> .	<i>Lineamientos para el examinador</i> , pág. 7	5 min	<input type="checkbox"/>
6. Pedir a los alumnos que completen los datos sobre sí mismos en la portada de la prueba.	<i>Lineamientos para el examinador</i> , pág. 9	5 min	<input type="checkbox"/>
7. Verificar que cada estudiante haya completado los datos sobre sí mismos requeridos en la portada.		10 min	<input type="checkbox"/>
8. Seguir las instrucciones para la Sesión 1.	<i>Lineamientos para el examinador</i> , pág. 11-13	60 min	<input type="checkbox"/>
9. Para los descansos, pedir a los estudiantes que salgan de la sala en fila y que dejen los exámenes sobre sus escritorios.		15 min	<input type="checkbox"/>
10. Seguir las instrucciones para la Sesión 2.	<i>Lineamientos para el examinador</i> , pág. 15-17	60 min	<input type="checkbox"/>
11. Para los descansos, pedir a los estudiantes que salgan de la sala en fila y que dejen los exámenes sobre sus escritorios.		15 min	<input type="checkbox"/>
12. Seguir las instrucciones para la Sesión 3.	<i>Lineamientos para el examinador</i> , pág. 19-21	70 min	<input type="checkbox"/>
13. Recoger todos los cuadernillos de examen y marcar su devolución mediante el formulario STBA.	Formulario STBA	10 min	<input type="checkbox"/>
14. Contabilizar todas las pruebas y asegurarse de que se hayan devuelto todas.	Formulario STBA	5 min	<input type="checkbox"/>
15. Despedir a la clase.		2 min	<input type="checkbox"/>

(continúa)

RECUADRO 13.1 (continúa)

Nombre		Fecha	
Tarea	Referencia	Hora	Completado
16. Firmar el formulario STBA.	Formulario STBA	2 min	<input type="checkbox"/>
17. Recoger y guardar todos los materiales de la prueba en la caja proporcionada, entre ellos los siguientes: i. Formulario STBA ii. Cuestionario para docentes iii. Formulario de observaciones del docente iv. Todas las pruebas completadas v. Todas las pruebas sin utilizar		10 min	<input type="checkbox"/>
18. Guardar los materiales de manera segura.		10 min	<input type="checkbox"/>
19. Devolver los materiales a su supervisor superior de distrito (SSD) para la Evaluación regional de matemáticas, ciencias e inglés (ERMCI).	Formulario de distribución de ERMCI del SSD	Tiempo de traslado	<input type="checkbox"/>
20. Devolver esta lista de verificación completa a su SSD.	Lista de verificación administrativa de ERMCI	2 min	<input type="checkbox"/>
Firma del examinador _____			

Fuente: Departamento de Educación de Filipinas 2004.



INFORMACIÓN A LAS ESCUELAS SOBRE LA EVALUACIÓN NACIONAL

Es necesario que los estudiantes estén motivados a esforzarse en la evaluación nacional. A los estudiantes generalmente se los motiva más cuando los docentes les explican el propósito de la prueba y cuando se garantiza que entiendan que los resultados se utilizarán para ayudar a mejorar la enseñanza y no para juzgarlos.

Todos los estudiantes necesitan sentirse animados a participar, especialmente aquellos con menos capacidades. La decisión sobre cuándo informar a los estudiantes sobre las fechas de la prueba dependerá de las circunstancias de la escuela. Si los estudiantes se sienten atemorizados por la prueba y eso hará que dejen de asistir a la escuela, entonces es preferible no decirles con exactitud cuándo se administrará. Si se sienten emocionados ante la posibilidad de realizar la prueba y es más probable que concurran a la escuela, entonces es preferible decirles cuándo se administrará.

La agencia que la implementa debe garantizar que las escuelas estén informadas sobre el propósito de la prueba con antelación a su administración. Se puede brindar la información mediante seminarios, cartas o teléfono. Es recomendable ser sincero y claro sobre los datos que se recopilarán, cómo se reportarán y utilizarán, y qué retroalimentación (si la hubiera) recibirá la escuela acerca del desempeño de los estudiantes.

Se informará a los directores y a los docentes de las escuelas que participan que sus escuelas y sus clases han sido seleccionadas para ayudar a aumentar la información sobre lo que los estudiantes saben y lo que no. Esta información se recopila para ayudar a mejorar el sistema de educación nacional. En las evaluaciones nacionales realizadas a una muestra no se juzga a las escuelas o clases. También se debe informar a los directores y a los docentes de las escuelas que participan que todos los datos de los exámenes y las respuestas a los cuestionarios serán confidenciales.

Es necesario contar con procedimientos que garanticen que las escuelas hayan acordado participar en la evaluación antes de que se asignen examinadores externos a cada una de ellas. También es necesario garantizar que los examinadores externos cuenten con las presentaciones necesarias, tales como una carta de aval, para asegurarse de que la escuela les brindará el apoyo necesario para que cumplan con sus responsabilidades. El volumen 3, *Implementación de una evaluación nacional del rendimiento académico*, contiene sugerencias adicionales sobre cómo informar a las escuelas e incluye el borrador de una carta.



GLOSARIO

análisis clásico de ítems: Una metodología estadística para el análisis de los datos de la prueba.

análisis de datos: Uso de una metodología estadística para analizar e interpretar los datos de la prueba.

analista de datos: La persona responsable del análisis estadístico de los datos.

área de aprendizaje: Un eje principal en el plan de estudios, tal como matemáticas o ciencias.

centro de corrección: El lugar donde se lleva a cabo la puntuación manual de los ítems de la prueba y donde se capacita y se supervisa a los correctores.

clave de opción múltiple: La opción correcta en un ítem de opción múltiple.

cohorte completa: Todos los estudiantes en un país que cumplen con ciertos criterios, tales como estar en un determinado año académico en un momento dado.

comité de revisión de ítems: Un grupo pequeño compuesto por entre tres y seis personas que de manera crítica revisan y mejoran todos los aspectos de los ítems para garantizar que sean de excelente calidad.

condiciones estandarizadas: Condiciones de prueba que se especifican en el manual de administración y que permanecen iguales para todos los estudiantes a quienes se les administra la prueba. Todos los estudiantes reciben la misma cantidad de apoyo, se les dan las mismas instrucciones y tienen la misma cantidad de tiempo para realizar la prueba.

confiabilidad de la prueba: El grado en el que la evidencia recopilada es suficiente para realizar generalizaciones.

conjunto de ítems: Un conjunto de ítems evaluados en una prueba de campo o prueba previa y de ítems seguros de pruebas anteriores que son apropiados para su uso en exámenes futuros.

corrección: Una revisión detallada de todos los aspectos de un texto para garantizar que sea claro, coherente y que no tenga errores.

corrector: Una persona que otorga puntajes manualmente a los ítems según una guía de puntuación.

corrector jefe: La persona responsable de la dirección del centro de corrección y de la resolución de las discrepancias de puntaje.

corrector principal: Un corrector experimentado responsable del cotejo de la puntuación manual para garantizar la coherencia y confiabilidad del trabajo de los correctores.

correlación punto biserial: Método utilizado en el análisis de los ítems para brindar una medición de la correlación (relación) entre el puntaje (correcto o incorrecto) que los estudiantes obtienen por un ítem individual y el puntaje total que obtienen en el resto de los ítems.

datos: Información recopilada de una prueba, generalmente ingresada en un programa de computadora.

datos agregados: Combinación de datos para dar un puntaje total, tal como un puntaje individual a partir de una prueba de 30 ítems.

distractores: Las opciones incorrectas en un ítem de opción múltiple.

doble evaluación: Un proceso en el que se califican dos veces las respuestas a los ítems generadas por los estudiantes. El segundo corrector no ve el primer puntaje.

ensayo: Una respuesta escrita extensa a una instrucción, generalmente de media página o más.

ensayo de prueba: Otro nombre para una “prueba previa” o “prueba de campo” que se lleva a cabo antes de la prueba final con una pequeña muestra de estudiantes, con el fin de establecer la calidad y la pertinencia de los ítems, los cuestionarios y los manuales de administración.

enunciado: La parte de un ítem de opción múltiple que antecede a las opciones, generalmente una pregunta, una oración incompleta o una indicación.

escala de grises: Un sombreado de impresión de color gris suave.

especificaciones técnicas: Especificaciones sobre los criterios que deben cumplir los ítems de la prueba final, entre ellos la proporción de ítems que aborda cada aspecto de un área curricular, la duración de la prueba, el formato de los ítems y cualquier otro criterio o limitación respecto del desarrollo de la prueba.

evaluación por cotejo: Una revisión del puntaje manual para controlar que coincida de manera sistemática con las categorías de puntaje de la guía de corrección. Por lo general, la verificación la realiza el corrector principal en el lugar para ofrecer retroalimentación inmediata a los correctores sobre la calidad de su trabajo.

evaluación de discrepancias: Procedimiento para resolver los puntajes contradictorios que surgen del cotejo o de la doble evaluación de las respuestas generadas por el estudiante. Por lo general, estos procedimientos son gestionados por el corrector jefe.

examinador: Una persona que supervisa la realización de la prueba y es responsable de garantizar que las condiciones de la prueba están estandarizadas según el manual de administración.

formulario final: Los cuadernillos de prueba que se administran a una población de muestra.

guías de corrección: Descripción de las categorías de puntuación que se utilizan para clasificar las respuestas a los ítems generadas por los estudiantes.

hoja de respuestas: Una hoja separada del cuadernillo de prueba que utilizan los estudiantes para registrar sus respuestas a los ítems de la prueba.

ítem: Una parte única de una prueba con un puntaje individual. Puede ser una pregunta, una oración sin terminar o una parte única de una prueba o cuestionario con un puntaje o código individual.

ítem de crédito parcial: Un ítem que tiene una o más categorías de respuesta correcta. Estas categorías generalmente son jerárquicas para los ítems en el formulario final de la prueba, pero pueden no serlo para la prueba preliminar, de campo o para el ensayo de ítems.

ítems discriminantes: Ítems que diferencian los desempeños de los estudiantes de gran capacidad y de poca: es decir, es más probable que los estudiantes que tienen gran capacidad respondan el ítem correctamente que los que tienen poca.

ítems de enlace: Ítems que se replican en dos o más cuadernillos de prueba para permitir que se comparen los datos de la evaluación de los cuadernillos.

ítems no sesgados: Ítems que examinan el rendimiento de los estudiantes de manera justa y no favorecen a ciertos grupos según características que no tienen importancia para el conocimiento o las destrezas que se evalúan.

ítems de opción múltiple: Ítems que requieren que los estudiantes seleccionen la única respuesta correcta para una pregunta a partir de una serie de opciones.

ítems de respuesta corta abierta: Ítems que requieren que un estudiante genere una respuesta breve, tal como una o dos oraciones, o varias modificaciones a una tabla, un cuadro o un diagrama.

ítems seguros: Ítems que se han mantenido fuera del dominio público. Es posible que hayan sido administrados en una prueba anterior, pero las condiciones de esa prueba han impedido su duplicación o divulgación.

manual de administración: Un conjunto escrito de instrucciones acerca de cómo, cuándo y dónde deben llevarse a cabo las pruebas.

El manual también puede incluir información sobre el traslado de los cuadernillos de prueba dentro y fuera de las escuelas.

marco: Un documento que define el propósito de la prueba e indica qué se debe medir, cómo debe medirse, por qué se mide y cómo debe reportarse.

material de estímulo: Textos, diagramas o cuadros que brindan el contexto para uno o más ítems.

muestra aleatoria: Un grupo de estudiantes seleccionado estadísticamente que cumple con ciertos criterios, entre ellos la distribución de variables clave que coincide con la distribución de las mismas variables en la cohorte completa.

objetividad de la prueba: El grado en el que la prueba no se ve alterada por la elección de la tarea o del asesor; es decir, la tarea es justa, completa y tiene criterios claros para determinar la puntuación.

población objeto de la prueba: Los estudiantes a quienes se administrará la prueba.

preguntas de respuesta cerrada: Ítems que requieren que los estudiantes generen una respuesta breve con un pequeño conjunto limitado de respuestas correctas.

prueba: Uno o más ítems a los que los estudiantes responden en condiciones estandarizadas. Los ítems se diseñan para permitir que los estudiantes demuestren sus conocimientos, destrezas y entendimiento.

prueba de campo: Otro nombre para un “ensayo de prueba” que se lleva a cabo antes de la prueba final con una pequeña muestra de estudiantes, con el fin de establecer la calidad y la pertinencia de los ítems y de los manuales de administración.

prueba piloto: Otro nombre para un “ensayo de prueba” que se lleva a cabo antes de la prueba final con una pequeña muestra de estudiantes, con el fin de establecer la calidad y la pertinencia de los ítems, los cuestionarios y los manuales de administración.

prueba previa: Otro nombre para un “ensayo de prueba” que se lleva a cabo antes de la prueba final con una pequeña muestra de alumnos,

con el fin de establecer la calidad y la pertinencia de los ítems, los cuestionarios y los manuales de administración.

puntaje: Puntos que se asignan a la respuesta de un estudiante según las categorías de una guía de puntuación.

puntaje dicotómico: Un ítem que se puntuó como correcto o incorrecto, de modo que tiene dos posibles puntajes: 0 y 1.

puntuación manual: La asignación de puntajes a las respuestas de los estudiantes a los ítems realizada por personas (no por máquinas) que otorgan puntajes.

responsables políticos: Funcionarios gubernamentales que dan forma a las políticas educativas.

respuestas faltantes: Ítems que el estudiante no ha intentado responder.

respuestas incorrectas: Ítems en los cuales la respuesta del estudiante encaja en la categoría de puntaje 0.

subcapítulo: Aspectos o agrupamientos dentro de las áreas de aprendizaje del plan de estudios; por ejemplo, es posible que matemáticas se encuentre dividida en subcapítulos de números, espacios, patrones y mediciones.

tabla de especificaciones: Otro nombre para las especificaciones técnicas.

unidad: Un grupo de ítems que se basa en una porción común de material de estímulo.

utilidad de la prueba: El grado en el que la prueba sirve a su propósito.

validez de la prueba: Un concepto amplio que implica hacer interpretaciones adecuadas y usar los puntajes o la información de la prueba de forma apropiada.

vinculación circular: La vinculación de un grupo de formularios de prueba, desde el primero hasta el último.

vinculación horizontal: La vinculación de los ítems entre formularios en el mismo nivel o año académico.

vinculación lineal: La vinculación de un grupo de formularios de prueba de uno al otro sin que el primer formulario se conecte nuevamente con el último.

vinculación longitudinal: La vinculación en el tiempo de los formularios de prueba.

vinculación vertical: La vinculación de los formularios de la prueba utilizados en distintos años académicos mediante la utilización de ítems comunes.



LECTURAS ADICIONALES

Allen, N. L., J. R. Donoghue, y T. L. Schoeps. 2001. Informe técnico de la evaluación NAEP 1998 Washington, DC: Centro Nacional de Estadísticas Educativas.

Baker, F. 2001. *The Basics of Item Response Theory*. College Park, MD: Centro de Información sobre Evaluaciones ERIC, Universidad de Maryland.

Beaton, A. E., y E. G. Johnson. 1989. "Overview of the Scaling Methodology used in the National Assessment." *Journal of Educational Measurement* 29: 163–75.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, y D. R. Krathwohl. 1956. *Taxonomía de objetivos educativos: Manual 1—Aspecto cognitivo*. Londres: Longmans, Green.

Campbell, J. R., D. L. Kelly, I. V. S. Mullis, M. O. Martin, y M. Sainsbury. 2001. *Marcos teóricos y especificaciones de evaluación de PIRLS 2001*. Chestnut Hill, MA: Centro de Estudios Internacionales, Boston College.

Centro de Investigación Educativa 1978. *Pruebas de evaluación Drumcora, Manual, Nivel II, Formulario A*. Dublín: Centro de Investigación Educativa.

Chatterji, M. 2003. *Designing and Using Tools for Educational Assessment*. Boston: Allyn y Bacon.

Consejo de Investigación Educativa de Australia. Sin fecha. *Literacy and Numeracy National Assessment (LANNA)*, preguntas de muestra,

Aritmética año 5. http://www.acer.edu.au/documents/LANNA_Y5NumeracyQuestions.pdf.

Consejo Directivo de Evaluaciones Nacionales. Sin fecha. *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: Departamento de Educación de EE. UU.

Departamento de Educación de Filipinas 2004. *Region-wide Assessment for Mathematics, Science, and English (RAMSE): Basic Education Assistance for Mindanao (BEAM)*. Manila: Departamento de Educación de Filipinas.

Departamento de Educación de Papúa Nueva Guinea 2003. *Cultural Mathematics Elementary Syllabus*. Port Moresby: Departamento de Educación de Papúa Nueva Guinea.

———. 2004. *National Curriculum Standards Monitoring Test*. Port Moresby: Departamento de Educación de Papúa Nueva Guinea.

Eivers, E., G. Shiel, R. Perkins y J. Cosgrove. 2005. Evaluación nacional de comprensión lectora en inglés, 2004. Dublín: Centro de Investigación Educativa.

Forster, M. 2000. *A Policy Maker's Guide to International Achievement Studies*. Camberwell, Australia: Consejo de Investigación Educativa de Australia.

———. 2000. *A Policy Maker's Guide to Systemwide Assessment Programs*. Camberwell, Australia: Consejo de Investigación Educativa de Australia.

Greaney, V., y S. B. Neuman. 1990. "The Functions of Reading: A Cross-Cultural Perspective." *Reading Research Quarterly* 25 (3):172–95.

Haladyna, T. M. 1999. *Developing and Validating Multiple-Choice Test Items*. 2.^a ed. Hillsdale, NJ: Lawrence Erlbaum.

Harlen, W., ed. 2008. *Student Assessment and Testing*. Vols. 1–4. Londres: Sage.

Howie, S. 1987. "Large-Scale Educational Assessment as Policy Research: Aspirations and Limitations." *European Journal of Psychology and Education* 2 (2): 157–65.

IEA (Asociación Internacional para la Evaluación del Rendimiento Educativo). 1998. *Third International Mathematics and Science Study—TIMSS Sample Items*. Chestnut Hill, MA: Centro de Estudios Internacionales, Boston College. <http://www.edinformatics.com/timss/pop1/mpop1.htm>, <http://timss.bc.edu/timss1995i/TIMSSPDF/BSIItems.pdf/>, and http://www.ed.gov/inits/Math/timss4_8.html.

- . 2007. *TIMSS 2003, Science Items, Released Set. Fourth Grade*. S011026. Chestnut Hill, Mass.: Centro de Estudios Internacionales TIMSS y PIRLS, Boston College. timss.bc.edu/PDF/T03_RELEASED_S4.pdf.
- Kirsch, I. 2001. *The International Adult Literacy Survey (IALS): Understanding What Was Measured*. Informe de investigación RR-01-25. Princeton, NJ: Servicio de Pruebas Educativas (ETS).
- Kubiszyn, T., y G. Borich. 2000. *Educational Testing and Measurement*. Nueva York: Wiley.
- Linn, R. L., y M. D. Miller. 2004. *Measurement and Assessment in Teaching: Student Exercise Manual*. Upper Saddle River, NJ: Prentice Hall.
- Linn, R. L., y S. B. Dunbar. 1992. "Issues in the Design and Reporting of the National Assessment of Educational Progress." *Journal of Educational Measurement* 29 (2): 177–94.
- Messick, S. 1987. "Large-Scale Educational Assessment as Policy Research: Aspirations and Limitations." *European Journal of Psychology and Education* 2 (2): 157–65
- . 1989. "Validity." En *Educational Measurement*, 3.^a ed., ed. R. L. Linn, 13–103. New York: American Council on Education/MacMillan.
- Ministerio de Educación de Nueva Zelanda 2002. *English in the New Zealand Curriculum*. Wellington: Medios de Aprendizaje para el Ministerio de Educación de Nueva Zelanda.
- Mullis, I. V. S., A. M. Kennedy, M. O. Martin, y M. Sainsbury. 2006. *Marcos teóricos y especificaciones de evaluación: Estudio Internacional de Progreso en Comprensión Lectora* 2.^a ed. Chestnut Hill, MA: Centro de Estudios Internacionales TIMSS y PIRLS, Boston College.
- Mullis, I. V. S., M. O. Martin, E. J. Gonzalez, y S. J. Chrostowski. 2004. *Informe internacional de matemática TIMSS 2003: Conclusiones de la evaluación del Estudio Internacional de Tendencias en Matemáticas y Ciencias de la IEA para cuarto y octavo grados*. Chestnut Hill, MA: Centro de Estudios Internacionales TIMSS y PIRLS, Boston College.
- Mullis, I. V. S., M. O. Martin, E. J. Gonzalez, K. D. Gregory, R. A. Garden, K. M. O'Connor, S. J. Chrostowski, y T. A. Smith. 2000. *Informe internacional de matemática TIMSS 1999. Conclusiones de la Repetición del Tercer Estudio Internacional de Matemáticas y Ciencias de IEA para octavo grado*. Chestnut Hill, Mass.: Centro de Estudios Internacionales, Boston College. timssandpirls.bc.edu/timss1999i/pdf/T99i_Math_2.pdf.

Nitko, A. J. 2004. *Educational Assessment of Students*. 4ta ed. Upper Saddle River, NJ: Pearson, Merrill, Prentice Hall.

PISA (Programa para la Evaluación Internacional de Alumnos). 2004. *Aprender para el mundo del mañana: Primeros resultados de PISA 200*. París: Organización para la Cooperación y el Desarrollo Económicos.



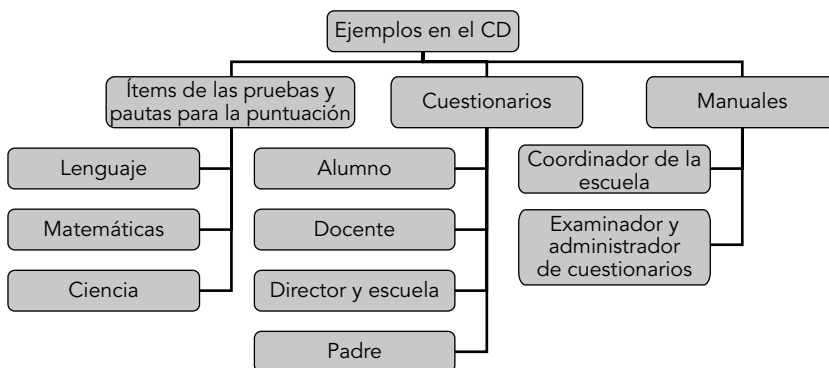
EJEMPLOS DE ÍTEMS DE PRUEBAS, ÍTEMS DE CUESTIONARIOS Y MANUALES DE ADMINISTRACIÓN EN EL CD

El disco compacto (CD) que acompaña a este libro contiene ejemplos de ítems de pruebas de rendimiento académico, guías para la evaluación y puntuación de pruebas, ítems de cuestionarios y manuales que han sido utilizados en diversos contextos, incluyendo evaluaciones nacionales e internacionales. La figura C.1 muestra la organización de los contenidos en el CD. La mayoría de los ítems, cuestionarios y manuales han sido publicados y se puede acceder a ellos en Internet. Se han reproducido en un CD para asistir a los equipos de evaluaciones pertenecientes a países en los que el acceso a Internet puede presentar dificultades. Agradecemos el apoyo de los editores y las organizaciones que autorizaron la reproducción de sus materiales originales en este CD, y los incluimos en una lista al final de este apéndice.

Los ejemplos que se incluyen en el CD ofrecen a los equipos a cargo de las evaluaciones nacionales ideas sobre los siguientes aspectos: tipos y formatos de ítems, guías para la evaluación y puntuación de pruebas, cobertura del contenido del currículo, estructura de las pruebas y los cuestionarios, y tipo de información publicada en los manuales de administración de las pruebas. Los equipos de evaluación nacional pueden utilizar esta información para diseñar sus propios instrumentos de evaluación, sus guías para

FIGURA C.1

Guía para acceder a los materiales en el CD sobre pruebas, cuestionarios y manuales



Nota: Hacer clic sobre el archivo "fuentes" para acceder a la fuente de ítems, cuestionarios y manuales individuales publicados, y a una lista de direcciones de sitios web (en caso de que corresponda) de los cuales se obtuvo dicha información.

la evaluación y puntuación de pruebas, y sus manuales. Deben tener en cuenta el currículo, la pertinencia del vocabulario y el formato de las pruebas a la hora de seleccionar o adaptar el material.

ÍTEMS DE PRUEBAS DE RENDIMIENTO ACADÉMICO

El CD contiene ítems de pruebas de matemáticas, lenguaje y ciencia. Esperamos que resulten útiles para los redactores de ítems de esas áreas, pues deberán desarrollar instrumentos de evaluación basados en su currículo nacional. No pretendemos que los equipos de evaluación nacional copien estos ítems. Dentro de cada materia, en primer lugar, se presentan archivos con ítems para grados de educación primaria, seguidos por archivos con ítems para el nivel de educación posprimaria y, finalmente, archivos con ítems que cubren tanto los niveles de educación primaria como los de posprimaria.

El CD recopila ítems extraídos de diferentes estudios nacionales de matemáticas, comprensión lectora y escritura, realizados en los Estados Unidos con alumnos de cuarto, octavo y décimo segundo grado, y de diferentes estudios con alumnos de 9, 13 y 17 años. Asimismo, contiene ítems de pruebas de matemáticas de cuarto grado

realizadas en el estado de Massachusetts e ítems utilizados en estudios nacionales realizados en Australia e Irlanda. El CD también presenta ítems publicados por tres programas internacionales de evaluación: el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) que evalúa el rendimiento en matemáticas y ciencias de alumnos de tercer, cuarto, séptimo y octavo grado y el último año de educación posprimaria; el Estudio sobre el Progreso Internacional de la Competencia en Lectura (PIRLS) que evalúa el lenguaje en alumnos de 4 grado; y el Programa para la Evaluación Internacional de Alumnos (PISA) que evalúa el desempeño de los alumnos de 15 años en lenguaje, matemáticas y ciencia.

Algunos de los ítems de las pruebas de lenguaje se aplican a textos extensos, un formato que puede no ser el adecuado en algunas evaluaciones nacionales. En ciertos casos, el formato del ítem para descargar es diferente al formato utilizado en el cuadernillo de prueba original. Es importante tener en cuenta que algunos de los ítems fueron diseñados para evaluar dos o más niveles de grado escolar.

El material incluido en el CD también contiene guías para la evaluación y puntuación de pruebas específicas.

CUESTIONARIOS

El CD contiene ejemplos de cuestionarios individuales para alumnos, maestros, padres, escuelas y directores. La mayoría de los cuestionarios han sido utilizados en estudios internacionales en países industrializados. Varios de ellos pertenecen a contextos educativos y escolares específicos. Los equipos de evaluación nacional deben considerar la posibilidad de adaptar algunas de las preguntas más relevantes para que estas reflejen la realidad económica, social y escolar de sus países.

MANUALES

El CD incluye manuales que brindan instrucciones específicas sobre cómo administrar las pruebas y los cuestionarios. También contiene manuales que describen los roles y las responsabilidades de quienes

coordinan las evaluaciones dentro de las escuelas. Entre las responsabilidades se incluyen las tareas que deben realizarse antes, durante y después de la administración de las pruebas y los cuestionarios. Los ejemplos abarcan temas tales como la preparación de los examinadores; la enumeración de materiales e insumos adecuados (por ejemplo, pruebas, cuestionarios, lápices y un reloj); la disposición de los asientos para minimizar el riesgo de que los alumnos se copien; el tratamiento de los ítems de ejemplo; la implementación de pautas de tiempo; y la identificación de tareas que debe realizar el examinador una vez terminada cada sesión.

Algunos materiales tendrán más relevancia en algunos países que en otros. Por ejemplo, algunos manuales hacen referencia a la corrección electrónica de pruebas o a los formularios de respuestas, que no suelen ser utilizados en las evaluaciones nacionales en los países en desarrollo. Se advierte a los usuarios que no deben replicar el contenido de los manuales, sino que deben seleccionar ideas que se adapten a sus contextos nacionales en particular. Los ejemplos se incluyen con el fin de ayudar a los equipos de evaluación nacional a desarrollar manuales basados en sus propias pruebas. Algunos manuales incluyen sugerencias sobre la selección de muestras de alumnos dentro de las escuelas.

AGRADECIMIENTOS

El Centro Nacional para Estadísticas de la Educación del Departamento de Educación de EE. UU. (<http://nces.ed.gov/nationsreportcard/about>) ha autorizado la reproducción de los ítems de pruebas, los manuales para examinadores y los cuestionarios de la Evaluación Nacional del Progreso Educativo.

La Asociación Internacional para la Evaluación del Rendimiento Educativo (<http://www.iea.nl/> y <http://timss.bc.edu/>) ha autorizado la reproducción de los ítems, los cuestionarios y los manuales para examinadores y coordinadores de escuelas de TIMSS y PIRLS.

La Organización para la Cooperación y el Desarrollo Económicos (<http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf>) ha autorizado

la reproducción de los ítems, los cuestionarios y los manuales para examinadores y coordinadores de escuelas de PISA.

El CD contiene ítems de matemáticas que han sido publicados por el Departamento de Educación de Massachusetts y que se encuentran disponibles en el sitio web del departamento en <http://www.doe.mass.edu/mcas/testitems.html>.

El Consejo Australiano de Investigación Educativa ha autorizado la reproducción de ejemplos de ítems y guías para la evaluación y puntuación de su Evaluación Nacional de Lectoescritura y Aritmética Elemental, comprensión lectora en tercer, quinto y séptimo año (<http://www.acer.edu.au/lanna/>).

El Centro de Investigación Educativa, Dublín (<http://www.erc.ie/index.php?s=7>), ha autorizado la reproducción de ítems de inglés y matemáticas, un cuestionario y un manual para examinadores.

El CD contiene un archivo que enumera las fuentes de los ítems, cuestionarios o manuales publicados y una lista de direcciones de sitios web (en caso de que corresponda) de los cuales se obtuvo dicha información. Pulse sobre el archivo “Fuentes” en el CD para acceder a esa información.

ECOAUDITORÍA

Declaración de beneficios medioambientales

El Grupo Banco Mundial tiene el compromiso de reducir su huella ambiental. En apoyo a dicho compromiso, la División de Publicaciones y Conocimiento impulsa las opciones de edición electrónica y la tecnología de impresión por encargo, desde centros regionales distribuidos por todo el mundo. En conjunto, estas iniciativas permiten reducir las tiradas y las distancias de envío, lo que redundará en un menor consumo de papel, menor uso de productos químicos, menores emisiones de gases de efecto invernadero y menor cantidad de residuos.

La División de Publicaciones y Conocimiento sigue las normas recomendadas sobre el uso de papel establecidas por la Green Press Initiative (Iniciativa de Prensa Ecológica). La mayor parte de nuestros libros se imprime con papel certificado por el Consejo de Administración de Bosques (FSC), y el contenido en papel reciclado de casi todos ellos oscila entre el 50 y el 100 por ciento. La fibra reciclada del papel de nuestros libros es o bien sin blanquear o blanqueada mediante procesos totalmente libres de cloro (TCF), procesos de fabricación sin cloro (PCF) o procesos de blanqueo libre de cloro elemental mejorado (EECF).

Puede encontrarse más información sobre la filosofía ambiental del Banco en <http://www.worldbank.org/en/about/what-we-do/crinfo>.



Evaluaciones nacionales del rendimiento académico

La evaluación efectiva del desempeño de los sistemas educativos es un componente clave en la formulación de políticas para optimizar el desarrollo del capital humano en todo el mundo. Los cinco libros de la serie *Evaluaciones nacionales del rendimiento académico* presentan conceptos clave de las evaluaciones nacionales de los niveles de rendimiento estudiantil, desde las cuestiones normativas que deben abordarse cuando se diseña y se lleva a cabo las evaluaciones hasta el desarrollo de las pruebas, el muestreo, la depuración de datos, las estadísticas, la redacción de informes y el uso de los resultados para mejorar la calidad de la educación.

Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico es el segundo de esta serie, diseñada para ayudar a desarrollar la capacidad de llevar a cabo evaluaciones técnicamente adecuadas de los niveles nacionales de rendimiento estudiantil. Presenta a los lectores las actividades ligadas al desarrollo de las pruebas de rendimiento académico, e incluye el desarrollo de un marco de evaluación, la redacción de ítems de tipo opción múltiple y respuesta construida, la realización de pruebas previas, la producción de cuadernillos de prueba y la calificación manual de los ítems.

Una parte dedicada a la elaboración de los cuestionarios describe el diseño de los mismos, la redacción de las preguntas, la codificación de las respuestas y la vinculación del cuestionario con los datos de puntuación de la prueba. La parte final expone el desarrollo de un manual de administración de la prueba, la selección de administradores de las pruebas y los contactos con las escuelas muestreadas.

El CD de acompañamiento contiene ejemplos de ítems procedentes de pruebas nacionales e internacionales, cuestionarios de muestra y manuales administrativos.

El libro debe servir de recurso útil para todos aquellos interesados en las evaluaciones educativas, especialmente el personal dedicado a la realización de evaluaciones nacionales de rendimiento estudiantil.



GRUPO BANCO MUNDIAL

ISBN 978-1-4648-0745-9



SKU 210745