

# Atención a la diversidad en evaluaciones educativas externas

Muestra de prácticas internacionales

Guadalupe Ruiz Cuéllar  
María Guadalupe Pérez Martínez  
Patricia Langford de la Rosa  
Adán Moisés García Medina



Cuadernos  
de investigación

43

**INEE**

Instituto Nacional para la  
Evaluación de la Educación  
México





# Atención a la diversidad en evaluaciones educativas externas

Muestra de prácticas internacionales

Guadalupe Ruiz Cuéllar  
María Guadalupe Pérez Martínez  
Patricia Langford de la Rosa  
Adán Moisés García Medina

Cuaderno  
de investigación

43

*Atención a la diversidad en evaluaciones educativas externas*  
*Muestra de prácticas internacionales*

Primera edición, 2015

ISBN: 978-607-7675-71-6

Guadalupe Ruiz Cuéllar, María Guadalupe Pérez Martínez,  
Patricia Langford de la Rosa y Adán Moisés García Medina

D.R. © Instituto Nacional para la Evaluación de la Educación  
Barranca del Muerto 341, Col. San José Insurgentes,  
Del. Benito Juárez; C.P. 03900 México, D.F.

Editora

María Norma Orduña Chávez

Corrección de estilo

Hugo Soto de la Vega

Formación

Heidi Puon Sánchez

Impreso y hecho en México.

Distribución gratuita. Prohibida su venta.

Consulte el catálogo de publicaciones en línea: [www.inee.edu.mx](http://www.inee.edu.mx)

El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del INEE. Se autoriza su reproducción parcial o total por cualquier sistema mecánico o electrónico para fines no comerciales y citando la fuente de la siguiente manera:

Ruiz, G., Pérez, M., Langford, P., y García, A. (2015). *Atención a la diversidad en evaluaciones educativas externas. Muestra de prácticas internacionales*. México: INEE.



# Índice

<b>Presentación</b> .....	<b>7</b>
<b>Marco conceptual</b> .....	<b>11</b>
1. La justicia en el proceso evaluativo .....	13
1.1. Estándares relacionados con la justicia de las evaluaciones .....	14
1.2. La metodología del Diseño Universal de las Evaluaciones .....	16
1.3. Diversidad cultural: un elemento esencial de la validez de las evaluaciones.....	21
2. Relación de la accesibilidad de las pruebas con las oportunidades de aprendizaje de los alumnos y el currículum .....	26
<b>Prácticas internacionales de atención a la diversidad.</b>	
<b>Países</b> .....	<b>29</b>
<b>Chile</b> .....	<b>30</b>
1.1. Descripción general .....	30
1.2. Evaluaciones de logro del SIMCE .....	31
1.3. Prácticas destacables de atención a la diversidad .....	32
1.4. Opiniones en los medios .....	42
<b>Estados Unidos de Norteamérica</b> .....	<b>43</b>
2.1. Descripción del Sistema Educativo Norteamericano .....	43
2.2. <i>No Child Left Behind Act</i> .....	43
2.2.1. Atención a la diversidad en las evaluaciones .....	45
2.2.2. Evaluación educativa de los alumnos con discapacidad .....	45
2.2.3. ¿Qué son los servicios especiales (acomodaciones)? .....	46
2.2.4. Acomodaciones vs. modificaciones .....	46
2.2.5. Acomodaciones para evaluación frente a acomodaciones para enseñanza .....	46
2.2.6. Acomodaciones o servicios especiales estándar frente a no estándar .....	46
2.2.7. Selección de las acomodaciones .....	47
2.2.8. Estudios de validez .....	49
2.3. National Center for Education Statistics (NCES) .....	49
2.3.1. Reporte de los resultados .....	51
2.3.2. Monitoreo.....	52
2.4. Educational Testing Service (ETS) .....	52
2.5. Opiniones en los medios .....	55

<b>Inglaterra .....</b>	<b>56</b>
3.1. Introducción .....	56
3.2. Descripción general de la evaluación o evaluaciones .....	57
3.3. Prácticas de atención a la diversidad .....	59
3.3.1. Conceptualización de la diversidad .....	59
3.3.2. Tipos de diversidad que considera .....	59
3.3.3. Atención a la diversidad en las diferentes fases del proceso evaluativo .....	60
3.3.3.1. En el diseño y aplicación de los instrumentos de evaluación .....	60
3.3.3.2. Atención de niños con Necesidades Educativas Especiales .....	63
Niños con debilidad visual .....	64
Niños con limitaciones auditivas .....	65
Lesiones o enfermedades .....	66
Niños con dificultades para concentrarse .....	66
3.3.3.3. Características de recursos de apoyo específicos .....	66
Modificaciones en cuanto al uso del tiempo .....	67
Incorporación de personal de apoyo en la aplicación .....	68
Otras modificaciones para facilitar el acceso a la prueba .....	70
3.3.3.4 Consideraciones adicionales para facilitar el acceso a las pruebas .....	71
3.4. Opinión pública .....	71
3.5. Estudios evaluativos sobre las Evaluaciones Nacionales .....	72
<b>Australia .....</b>	<b>73</b>
4.1. Descripción del Sistema Educativo .....	73
4.1.1. Australian Qualifications Framework (AQF) .....	74
4.2. Sistema de evaluación educativo .....	75
4.2.1. Programa Anual de Evaluación Nacional en Alfabetización y Aritmética (NAPLAN) .....	75
4.3. Prácticas de consideración de la diversidad .....	76
4.4. Opinión pública .....	78
<b>Prácticas internacionales de atención a la diversidad.</b>	
<b>Organismos internacionales .....</b>	<b>79</b>
<b>Programa Internacional de Evaluación de los Estudiantes (PISA) de la OCDE .....</b>	<b>80</b>
5.1. Descripción general y propósito .....	80
5.2. Marco conceptual .....	81
5.3. Marco contextual .....	82
5.4. Características de la prueba .....	82
5.5. Cobertura .....	83
5.6. Atención a la diversidad .....	84
<b>Oficina Regional de Educación de la UNESCO para América Latina y el Caribe,</b>	
<b>Santiago de Chile (OREALC/UNESCO Santiago) .....</b>	<b>86</b>
6.1. Descripción general .....	86
6.2. Marco conceptual en que se fundamenta .....	88
6.3. Proceso de diseño del SIRIED .....	91
6.4. Modelo de análisis para la operación del SIRIED .....	92
6.5. Primeros resultados de la implementación del Sistema .....	97

<b>Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).....</b>	<b>99</b>
7.1. Descripción general .....	99
7.2. Prácticas de atención a la diversidad.....	101
7.3. Opinión de los medios.....	103
7.4. Estudios sobre las evaluaciones SACMEQ.....	103
<b>Conclusiones.....</b>	<b>104</b>
<b>Referencias bibliográficas .....</b>	<b>113</b>

## Índice de tablas y figuras

### Tablas

<b>1</b> Acomodaciones empleadas con alumnos con necesidades educativas especiales o alguna discapacidad .....	14
<b>2</b> Orientaciones para la justicia de las evaluaciones del ETS.....	16
<b>3</b> Formato para la revisión del diseño universal de los ítems.....	20
<b>4</b> Acomodaciones para alumnos cuya primera lengua no es inglés ( <i>English Language Learners</i> ) de acuerdo al apoyo lingüístico que ofrecen .....	23
<b>5</b> Procesos comunes en las acomodaciones de traducción .....	24
<b>6</b> Perspectivas teóricas en las que se apoya el Smarter Balanced.....	25
<b>7</b> Principios para incrementar la accesibilidad de los reactivos del Smarter Balanced .....	26
<b>8</b> Preguntas relacionadas con la justicia y equidad de las evaluaciones.....	27
<b>9</b> Guía para elegir acomodaciones o servicios especiales.....	48
<b>10</b> Acomodaciones o servicios especiales generales.....	54
<b>11</b> Tipos de escuelas en Inglaterra y la participación que tienen en las pruebas Key Stage 2 .....	58
<b>12</b> Criterios comunes para juzgar el cumplimiento de los acuerdos de las evaluaciones .....	60
<b>13</b> Condiciones que deben cumplir las Evaluaciones Nacionales .....	61
<b>14</b> Características principales de los paradigmas de integración educativa y educación inclusiva .....	88
<b>15</b> Dimensiones y categorías del SIRIED .....	94
<b>16</b> Matriz de indicadores del SIRIED.....	95

### Figuras

<b>1</b> Proceso del diseño universal .....	18
<b>2</b> Proceso general de las Evaluaciones Nacionales .....	61

## DIRECTORIO

### JUNTA DE GOBIERNO

Sylvia Irene Schmelkes del Valle  
CONSEJERA PRESIDENTA

Eduardo Backhoff Escudero  
CONSEJERO

Teresa Bracho González  
CONSEJERA

Gilberto Ramón Guevara Niebla  
CONSEJERO

Margarita María Zorrilla Fierro  
CONSEJERA





# Presentación

**Este documento describe una muestra de prácticas internacionales en materia de consideración de la diversidad en evaluaciones educativas externas.** Responde a una solicitud hecha en agosto de 2014 por el Instituto Nacional para la Evaluación de la Educación (INEE) a Innovación y Desarrollo de la Educación, S. C. (IDE) para desarrollar un estudio que permitiera identificar y caracterizar este tipo de prácticas en evaluaciones externas realizadas tanto por países como por organismos internacionales.

El INEE sostiene que “la manera en que el sistema educativo se evalúa debe considerar la diversidad socioeconómica, cultural y el principio de equidad” (DGEOE, 2014). En este sentido, la revisión de prácticas internacionales de consideración a la diversidad puede constituir un insumo valioso para la formulación de “criterios técnicos para el desarrollo de instrumentos y metodologías de evaluación desde la perspectiva de la diversidad” (DGEOE, 2014).

Para realizar el estudio documental que se presenta en este informe se siguió un proceso que comprendió las siguientes cuatro etapas:

1. Identificación de prácticas internacionales de consideración de la diversidad en las evaluaciones externas de países y organismos internacionales que representen a los cinco continentes.
2. Caracterización de las prácticas internacionales de consideración de la diversidad.
3. Recopilación y análisis de estudios evaluativos de dichas prácticas y caracterización de las principales posturas sobre ellas en la opinión pública de los países respectivos o a nivel internacional.
4. Clasificación de los materiales empleados para integrar el documento.

La primera etapa inició con la exploración de los países y organismos internacionales que podrían ser contemplados en el estudio. De acuerdo con la solicitud hecha por el INEE, se buscó que en la selección de países estuvieran representados los cinco continentes, esperando que al final la muestra de prácticas incluyera un país de cada continente:

Región	Países
América del Norte	Canadá y Estados Unidos
Latinoamérica	Brasil, Chile, Perú, Paraguay y Bolivia
Europa	España e Inglaterra
África	Sudáfrica
Asia	India, Corea del Sur, China y Japón
Oceanía	Australia

Se revisaron las fuentes electrónicas disponibles en inglés, español y francés sobre los sistemas de evaluación de cada uno de los países para valorar la suficiencia de información sobre las evaluaciones y la presencia de prácticas en materia de consideración de la diversidad.

Las fuentes electrónicas consultadas fueron: páginas web de los sistemas educativos y de las pruebas o sistemas de evaluación existentes en los países estudiados, documentación específica sobre las evaluaciones realizadas, como informes y manuales, lo mismo que artículos de investigación y de prensa.

En la búsqueda de información se utilizaron diversos descriptores; por lo dicho antes, algunos referían a cada país o sistema de evaluación externa considerado. En cuanto a los descriptores sustantivos, uno de los principales fue el de “justicia de las evaluaciones” (*assessment fairness; testing fairness*); así, la búsqueda en cada país se hizo contemplando la combinación de este descriptor con el del país o sistema de evaluación de éste.<sup>1</sup> También se utilizaron descriptores asociados a los siguientes tipos de diversidad: género (*gender*), lengua (*first language, additional language*),<sup>2</sup> niños migrantes (*migrant children*), comunidades viajeras (*traveller communities*), minorías étnicas (*ethnic minorities*) y niños con discapacidades (*children with disabilities*). Además, para cada país y sistema de evaluación, se buscó información específica sobre validez cultural (*cultural validity*), estudios sobre el funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés), o estudios de sesgo (*bias*).

A partir de la valoración inicial de la información reunida sobre los países y organismos internacionales, y tomando en cuenta la suficiencia de información para realizar la caracterización de las prácticas de consideración de la diversidad, se eligieron los siguientes casos:

<b>Países</b>	Chile
	Estados Unidos
	Inglaterra
	Australia
<b>Organismos internacionales de evaluación</b>	Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA)
	Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC)
	Consortio para Monitorear la Calidad Educativa en el Sur y Este de África (SACMEQ)

Como se observa, Asia y África no se encuentran representados en la selección de países. Esto se debió fundamentalmente a la dificultad para encontrar información y documentación sobre las pruebas o sistemas de evaluación en los idiomas a los que tenía acceso el equipo a cargo del presente estudio. Sin embargo, en la selección de los organismos internacionales, se incluyó al Consorcio para Monitorear la Calidad Educativa en el Sur y Este de África (SACMEQ, por sus siglas en inglés) lo que implicó que sólo del continente asiático no fuera considerado algún país.

A partir de la decisión sobre los países y organismos internacionales por incluir en el estudio, se desarrolló la caracterización de las prácticas internacionales de consideración de la diversidad. En esta etapa se analizó la documentación reunida de acuerdo a un sistema de categorías y de acopio de información que incluyó la descripción de las pruebas o sistemas de evaluación, y de manera específica, las prácticas en materia de atención a la diversidad.

<sup>1</sup> En este caso, por poner un ejemplo, la búsqueda para el caso de Inglaterra y Chile, respectivamente, se hizo de la siguiente manera: *assessment fairness* y *key stage 2*, justicia de las evaluaciones y SIMCE.

<sup>2</sup> Estos casos se enfocaron en la atención de los grupos de niños cuya primera lengua es distinta de la que se utiliza en las pruebas (*additional language*).

Las categorías utilizadas para la descripción de los sistemas de evaluación fueron las siguientes:

- Propósito de la evaluación
- Objeto de la evaluación: logro educativo (precisando, según fuera el caso, asignatura, temas, contenidos evaluados), procesos educativos
- Alcance: subsistema, escuelas, alumnos
- Población a la que se dirige: niveles, grados
- Impacto: usos y consecuencias pretendidos de los resultados
- Antecedentes
- Fundamentos conceptuales del objeto de evaluación

En cuanto a la atención a la diversidad las categorías empleadas fueron:

- Conceptualización de la diversidad en el sistema o programa de evaluación
- Tipos de diversidad considerados
- Normas o criterios de consideración de la diversidad y procedimientos aplicados:
  - En el diseño de los instrumentos de evaluación
  - En la implementación de la evaluación (v. g., logística, administración de los instrumentos)
  - En el procesamiento y análisis de la información
  - En la interpretación de los resultados
  - En la presentación de los resultados
  - En la difusión de los resultados
  - En el uso de los resultados
- Recursos humanos y materiales requeridos para la consideración de la diversidad en los procesos a los que aplicará

La información anterior se registró en fichas de trabajo a partir de las cuales se elaboraron los apartados de este informe que sintetizan los resultados respecto a cada una de las categorías consideradas.

La tercera etapa, consistente en la recopilación y análisis de estudios evaluativos de las prácticas de consideración de la diversidad y la descripción de las principales posturas al respecto en la opinión pública de los países o a nivel internacional, implicó la búsqueda en dos tipos de fuentes principalmente: por un lado, aquellas de carácter académico que dieran cuenta de estudios sobre las prácticas de consideración de la diversidad en evaluaciones externas; y por otro, fuentes de amplia circulación como periódicos y revistas especializadas de divulgación, así como en periodos cercanos a la difusión de resultados.

Finalmente, los materiales identificados y empleados para integrar este documento fueron clasificados de acuerdo al país u organismo internacional al que se refieren, y fueron resguardados en dispositivos externos y en línea.

Considerando todos los elementos expuestos, se tomó la decisión de organizar este documento en cuatro apartados. En el primero se formula el marco conceptual que sirvió de base para realizar la caracterización de las prácticas evaluativas de consideración de la diversidad. Los capítulos segundo y tercero son los centrales del informe; en el segundo se incluyen los hallazgos relativos a los cuatro países cuyos sistemas de evaluación y monitoreo han incorporado elementos de consideración de la diversidad: Chile, Estados Unidos, Inglaterra y Australia. En el tercer apartado se hace lo propio para tres organismos internacionales que desarrollan programas de evaluación

educativa externa y que exhiben algunas prácticas de evaluación inclusiva: el Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA, por sus siglas en inglés), la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC), y el SACMEQ. En el cuarto apartado se presentan las conclusiones generales del informe.

Es importante señalar que al iniciar este estudio se tenía la expectativa de encontrar sistemas de evaluación y monitoreo que incluyeran prácticas para atender múltiples tipos de diversidad en las distintas fases del proceso evaluativo; sin embargo, uno de los hallazgos principales es que, en general, los sistemas estudiados se concentran en la atención de algún tipo de diversidad y sólo en una o algunas fases del proceso evaluativo, pero no en todas.

Las prácticas incluidas en este informe no deben verse como un modelo a seguir en materia de consideración de la diversidad en las evaluaciones externas que realiza el INEE; al menos no sin haber analizado cuidadosamente sus efectos, el contexto en el que han sido empleadas y la distancia (o cercanía) entre éste y el de nuestro propio sistema educativo. Su selección ilustra mecanismos y procedimientos que han sido empleados para: identificar grupos de estudiantes cuyas necesidades específicas requieren ser atendidas para facilitar su acceso a las pruebas estandarizadas, planificar la atención a estas necesidades, implementar medidas destinadas a ampliar el acceso a las pruebas, y evaluar la efectividad de esas medidas.

A partir de la revisión de prácticas de atención a la diversidad realizada, se puede afirmar que el proceso de atención de la diversidad en las evaluaciones estandarizadas es muy complejo y que por ende la toma de decisiones no es sencilla. Requiere que en principio, se analicen los grupos de alumnos que sustentan las pruebas, y se identifiquen aquellos subgrupos para quienes la prueba podría representar constructos distintos y llevar a puntajes que tuvieran diferentes significados. Posteriormente, implica que se analice el tipo de necesidades que tienen estos grupos y la manera en la cual podrían ser abordadas en los operativos de evaluación y particularmente en las pruebas. Finalmente, requiere del diseño de mecanismos de monitoreo continuo e investigación, para valorar la eficacia de las medidas que se implementen y sus condiciones de aplicación. Por lo anterior, se sostiene que las decisiones con respecto a la atención de la diversidad en las evaluaciones no pueden basarse en experiencias propias de otros lugares, con procesos, propósitos y condiciones específicas.

Una de las limitaciones de este estudio fue la disponibilidad de información sobre los sistemas de monitoreo y evaluación en forma electrónica, principal mecanismo de consulta. Es probable que si se incorporaran otras fuentes o procedimientos, como por ejemplo, entrevistas con agentes clave en cada uno de los sistemas educativos (autoridades educativas e investigadores especializados en evaluaciones de aprendizajes), podrían tenerse resultados enriquecidos. Otra limitación fueron los idiomas de acceso a la información del equipo de investigación, pues, como se señaló, la búsqueda se hizo en inglés, español y francés, lo cual constituyó una desventaja, principalmente al momento de hacer la búsqueda en países asiáticos.

Con todo, el equipo considera que el informe aporta elementos valiosos para enriquecer los procesos evaluativos que ocurren en el sistema educativo nacional, y sobre todo, para avanzar en la justicia con que se tienen que realizar las evaluaciones. Asimismo, se agradece el apoyo brindado por las autoridades del INEE, en particular de la Dirección General de Evaluación de la Oferta Educativa y, la Dirección de Evaluación y Diversidad, y los comentarios del Dr. Guillermo Solano-Flores a una versión previa de este documento.



Marco conceptual



### La preocupación por atender la diversidad en el marco de las evaluaciones educativas

surgió hasta la segunda mitad del siglo XX (Zieky, 2006), probablemente relacionada con la promoción de los derechos civiles de las personas. Actualmente se reconoce que la atención a la diversidad de las características de los sustentantes es un asunto fundamental de la validez (AERA, APA y NCME, 2014) y enmarca la *justicia*<sup>1</sup> de las evaluaciones.

El mayor desarrollo a este respecto se ha hecho en cuanto a las pruebas educativas y psicológicas, en las que “se obtiene una muestra de un comportamiento de un examinado en un dominio específico, y posteriormente se evalúa y califica usando procesos estandarizados” (AERA *et al.*, 2014: 2).

En los procesos de validación de estas pruebas se analiza si son justas para diferentes grupos, dadas sus cualidades y experiencias específicas; partiendo de reconocer que las evaluaciones son actividades sociales, “que sólo pueden ser entendidas considerando los contextos culturales, económicos y políticos en los que operan” (Gipps y Stobart, 2010: 56).

En este sentido, se aspira a que las evaluaciones ofrezcan oportunidades similares para que los sustentantes muestren su talento, y que los puntajes que se obtengan “no dependan de factores distintos a los que la prueba pretende medir” (Solano-Flores, 2011: 3). Por lo tanto, la justicia en las evaluaciones educativas y psicológicas se entiende como “la atención a las características individuales y contextos de las pruebas de tal manera que los puntajes lleven a interpretaciones válidas para los destinatarios [...] una prueba es justa, cuando refleja el mismo constructo para todos los sustentantes, y los puntajes tienen el mismo significado para todos los individuos de la población objetivo” (AERA *et al.*, 2014: 50).

La mayoría de la literatura revisada con respecto a la justicia de las evaluaciones muestra que la atención se concentra en el proceso evaluativo; esto es, se proponen guías, procesos y métodos que pueden orientar las diferentes etapas del proceso de tal manera que se maximice la *accesibilidad* de la prueba a los sustentantes, y se generen interpretaciones y usos válidos de los resultados. Las diferentes etapas a las que se alude comprenden: el diseño de los instrumentos de evaluación; la implementación de la evaluación (*v. gr.*, logística, administración de los instrumentos); el procesamiento y análisis de la información; así como la interpretación, presentación, difusión y uso de los resultados.

Además de este enfoque en el proceso evaluativo, se encontraron también contribuciones en las cuales la justicia de las evaluaciones se considera “inseparable de la justicia en las oportunidades de aprendizaje de los alumnos y la oferta del currículo” (Stobart, 2005: 278), pues,

<sup>1</sup> En la literatura internacional se emplea el término inglés *fairness*, que ha sido traducido como “justicia” o “equidad” (Stobart, 2005, 2009).

ambas median las respuestas de los sustentantes ante la medición de un constructo y pueden poner en riesgo la validez de las interpretaciones de los resultados, sus usos y consecuencias (AERA *et al.*, 2014). Es importante señalar que el desarrollo con respecto a este último enfoque es muy escaso.

De acuerdo con lo anterior, este capítulo se divide en dos apartados. El primero se enfoca en la justicia de las evaluaciones centrada en el proceso evaluativo; presenta algunos estándares que se han desarrollado para orientar la justicia de las evaluaciones, la metodología del diseño universal como una de las herramientas que puede ayudar a maximizar la accesibilidad de las pruebas, y las aportaciones para promover la atención a la diversidad cultural, en general, y lingüística, en particular, en las evaluaciones.

El segundo apartado, describe las aportaciones con respecto al vínculo de las oportunidades de aprendizaje y a la oferta curricular con la justicia de las evaluaciones. Es un apartado breve, debido al desarrollo que hay en este campo.

## 1. LA JUSTICIA EN EL PROCESO EVALUATIVO

En la literatura revisada se encuentra un consenso en aceptar que la justicia de las evaluaciones debe ser atendida en todas las etapas del proceso evaluativo desde su diseño (AERA *et al.*, 2014; Gipps y Stobart, 2010; Solano-Flores, 2011).

En este apartado se describen en primera instancia algunos estándares que se han desarrollado con respecto a la justicia de las evaluaciones, que tienen que ver con las diferentes etapas del proceso evaluativo. Posteriormente se describe la metodología del diseño universal como una de las herramientas que puede ser utilizada para maximizar la accesibilidad de las pruebas a los sustentantes a quienes se dirige y se presentan a detalle los métodos que se utilizan para la revisión de los ítems.

Las características de accesibilidad han sido clasificadas en tres tipos, de acuerdo con la proporción de estudiantes que se pueden beneficiar de ellas: herramientas universales (rasgos disponibles para todos los estudiantes), apoyos designados (características que están disponibles para estudiantes que han sido identificados con necesidades específicas por sus educadores) y acomodaciones (cambios en los procedimientos o materiales con los cuales se intenta incrementar la igualdad de acceso) (*Smarter Balanced Assessment Consortium*, citado por Solano-Flores *et al.*, 2014).

En la literatura revisada se precisan medidas con respecto a dos grandes grupos de características de los alumnos: culturales, que comprenden la diversidad lingüística, étnica, socioeconómica, de género, racial y religiosa, entre otras; y las referentes a las necesidades educativas especiales de los alumnos, que aluden por lo general a dificultades de aprendizaje (ritmos de aprendizaje significativamente distintos a los del grupo de edad del alumno) o alguna discapacidad.

Las acomodaciones empleadas con alumnos con necesidades educativas especiales o alguna discapacidad se presentan en la tabla 1:

**Tabla 1** Acomodaciones empleadas con alumnos con necesidades educativas especiales o alguna discapacidad

a) En la presentación de la prueba. Ajustes en la presentación visual (por ejemplo, impresiones ampliadas), auditiva (lectura de los ítems a los estudiantes usando una pantalla o texto) o multisensorial de la prueba.
b) En relación al registro de las respuestas y al proceso de responder la prueba. Tiene que ver con las modificaciones en las respuestas visuales (por ejemplo, permitir que los estudiantes marquen las respuestas en el cuadernillo de la prueba, o que señalen las respuestas, entre otros), y con la provisión de materiales o equipo para resolver u organizar respuestas (por ejemplo: reglas, materiales manipulativos para matemáticas, entre otros).
c) Prácticas relacionadas con la aplicación de la prueba. Incluyen condiciones con respecto al tiempo (por ejemplo: aplicar la prueba en múltiples sesiones; tener un programa de aplicación flexible, entre otros) y los espacios de la aplicación (por ejemplo: permitir que los estudiantes se muevan, administrar la prueba de manera individual o en pequeños grupos, hacer ajustes a la iluminación del salón donde se aplica la prueba, entre otros).
d) El personal que aplica la prueba. Esto comprende, por ejemplo, la participación de traductores de lenguaje de señas, para niños con deficiencias auditivas.
e) Motivación. Comprende prácticas que animen a los estudiantes a responder.
f) Preparación para la prueba. Incluye la enseñanza de habilidades para responder una prueba y práctica de ejercicios.

Fuente: Adaptado de Shafer Willner *et al.* 2008: 18.

Las acomodaciones de atención a las necesidades de los alumnos cuya primera lengua difiere de aquella en la que se aplica la prueba serán descritas a detalle en la tercera sección de este apartado. Aquí señalamos solamente que son regularmente clasificadas en apoyos directos e indirectos (Shafer Willner *et al.* 2008: 16).

Los apoyos directos tienen que ver con modificaciones o ajustes en el lenguaje, y pueden ser orales (por ejemplo: guión en lengua en que se aplica la prueba y emisión de respuestas de forma oral, entre otros) o escritos (por ejemplo: simplificación del idioma en que se aplica la prueba, provisión de recursos de consulta, traducción de la prueba, provisión de materiales en dos idiomas, y ajustes con respecto a la lengua en que los alumnos dan respuesta a la prueba).

Los indirectos tienen que ver con las condiciones de la aplicación (por ejemplo: una mayor cantidad de tiempo para dar respuesta a la prueba). De hecho, por lo general se utilizan de manera conjunta estos dos tipos de acomodaciones.

### 1.1. Estándares relacionados con la justicia de las evaluaciones

Para promover y valorar la justicia en las evaluaciones se han definido estándares que guían los diferentes procesos y acomodaciones que se incluyen en las evaluaciones estandarizadas.

En los estándares definidos por AERA *et al.* (2014), los relacionados con la justicia de las evaluaciones se dividen en cuatro grupos:

1. Diseño de la prueba, desarrollo, administración y procedimientos de calificación que minimicen las barreras para una interpretación válida del mayor rango posible de individuos y subgrupos relevantes.



En la definición puede verse que los estándares que se encuentran en este grupo tienen que ver con las diferentes fases por las que atraviesa una evaluación. Comprende, por ejemplo, la definición clara del constructo a medir y las características de los individuos y subgrupos a quienes se pretende aplicar las evaluaciones, la definición de consignas e ítems en los que los estímulos no necesarios sean erradicados de tal manera que se maximice el acceso a los sustentantes para manifestar su posición con respecto al constructo medido, la realización de estudios piloto en los que se incluyan subgrupos relevantes, y la provisión de un tratamiento comparable durante la administración y calificación de la prueba.

2. Validez de las interpretaciones de puntajes de las pruebas en relación con los usos propuestos para la población objetivo evaluada.

Se consideran estándares relacionados con la manera de analizar las diferencias entre subgrupos de la población evaluada (“contenido de la prueba, estructura interna de las respuestas, relación de los puntajes con otras variables, y procesos de respuesta de algunos individuos [...] estudios sobre la precisión y exactitud de los puntajes”, (AERA *et al.*, 2014: 65), la validez de criterio cuando se hagan predicciones basadas en los puntajes de las pruebas, y la validez de la interpretación de los puntajes, en especial cuando se consideran preguntas de respuesta construida.

3. Acomodaciones para remover las barreras irrelevantes y apoyar las interpretaciones válidas de los puntajes y sus usos propuestos.

Este grupo de estándares tiene que ver con los ajustes que se hacen para remover las barreras que son irrelevantes para la medición del constructo. Las adaptaciones pueden darse en cuanto al “lugar de la aplicación, presentación de la prueba, interfaz, requerimientos de respuesta, o añadir otros individuos al proceso de aplicación” (AERA *et al.*, 2014: 67). También comprende la documentación de los procedimientos para definir las adaptaciones, la difusión de la información a los sustentantes, y el análisis de la validez de las interpretaciones, entre otros.

4. Garantías en contra de interpretaciones inadecuadas de los puntajes de la prueba con respecto a los usos pretendidos.

Comprende cuatro estándares relacionados con la provisión de información que oriente la interpretación adecuada de los puntajes para los diferentes grupos de población que participaron en la prueba. Se pide que los desarrolladores de las pruebas provean evidencia de la aplicabilidad de la prueba para los diferentes grupos y “expliciten precauciones para evitar usos inadecuados de los resultados” (AERA *et al.*, 2014: 70). También se requiere que cuando los “resultados de investigaciones muestren que los puntajes de algunos subgrupos poblacionales fueron afectados diferencialmente por características no relacionadas con el constructo medido, [...] sólo se usen los resultados de los subgrupos para los cuales hay suficiente evidencia de validez para apoyar las interpretaciones de puntaje para los usos propuestos” (AERA *et al.*, 2014: 70).

En cuanto a la publicación de resultados, se señala que antes de agregar puntajes de diferentes subgrupos poblacionales se debe asegurar y proveer evidencia de su comparabilidad y, en caso contrario, señalar que los resultados no son comparables. Se advierte también que los

resultados de las pruebas no pueden utilizarse como único elemento para tomar decisiones con respecto a la colocación de alumnos en programas de educación especial; se deben utilizar múltiples fuentes de información.

En el *Educational Testing Service* (ETS), también se han desarrollado guías específicas para promover la justicia de las evaluaciones respecto a tres fuentes de varianza irrelevantes para el constructo medido: cognitiva, afectiva y física (ETS, 2009). En la tabla 2 se muestran las orientaciones vigentes.

**Tabla 2** Orientaciones para la justicia de las evaluaciones del ETS

Orientaciones	Descripción
Evitar fuentes cognitivas de varianza irrelevantes para el constructo medido	Cuando para responder un ítem se requieren habilidades o conocimientos no relacionados con el constructo medido.
Evitar fuentes afectivas de varianza irrelevantes para el constructo medido	Cuando se utiliza lenguaje o imágenes que provoquen emociones fuertes por parte de los sustentantes que interfieran con su habilidad para responder un ítem correctamente.
Evitar fuentes de varianza física irrelevantes para el constructo medido	Cuando las características de la prueba interfieren con la habilidad de los sustentantes para asistir a la aplicación, ver, oír o percibir los ítems o estímulos de la prueba.

Fuente: Elaboración propia a partir de ETS (2009).

## 1.2. La metodología del Diseño Universal de las Evaluaciones

El Diseño Universal de las Evaluaciones es una metodología que se ha impulsado para maximizar su accesibilidad. El principal objeto de la metodología del diseño es la *accesibilidad de las evaluaciones*, la cual se entiende como la oportunidad —sin obstáculos— que “debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir” (AERA *et al.*, 2014: 49). Se parte del supuesto de que el diseño de una prueba puede impedir el acceso a demostrar el desempeño de los estudiantes, por ejemplo, cuando el diseño tipográfico no permite que los sustentantes con alguna debilidad visual puedan leer los reactivos de una prueba que no está destinada a medir la capacidad visual.

El diseño universal es un concepto que proviene de la arquitectura (Johnstone, Thompson, Bottsford-Miller y Thurlow, 2008). En el campo de la arquitectura, el diseño universal postula que las acomodaciones para promover el acceso a diversos usuarios (como rampas, amplitud de puertas y pasillos, entre otros) deben considerarse desde la planificación del edificio, no cuando éste ya haya sido terminado (Johnstone *et al.*, 2008).

En el campo educativo, el diseño universal ha sido utilizado tanto en la enseñanza como en la evaluación. En este último campo el diseño universal se define como “aquellas evaluaciones que son diseñadas y desarrolladas desde el inicio para favorecer la participación del mayor número de estudiantes, y que los resultados lleven a inferencias válidas acerca del desempeño de todos los estudiantes” (Thompson, Johnston y Thurlow, 2002: 5).

El diseño universal pretende “promover el acceso a la mayor cantidad de sustentantes, sin importar su raza, origen étnico, género, estado socioeconómico, discapacidad, lengua o contexto cultural” (AERA *et al.*, 2014). Es importante señalar que el desarrollo del diseño universal de las

evaluaciones fue impulsado principalmente por los desafíos que representaba la medición de aprendizajes de estudiantes que se encuentran aprendiendo la lengua en la cual se aplican las evaluaciones<sup>2</sup> y quienes tienen alguna discapacidad temporal o permanente (Johnstone *et al.*, 2008). Sin embargo, en la actualidad se reconoce, que estudiantes que hablan la lengua en que se aplica la evaluación o que no cuenten con discapacidades también podrían enfrentar “dificultades con formatos o diseños que les resulten confusos” (Thompson *et al.*, 2002: 3)

De acuerdo con lo anterior, mediante el diseño universal se busca dar respuesta a las necesidades de los sustentantes que no interfieren con la medición del constructo, pero que, si no fueran tomadas en cuenta, les impedirían el acceso a la medición o limitarían la validez de las interpretaciones de los resultados. Por ejemplo: la impresión magnificada de una prueba de matemáticas para alumnos con algún tipo de debilidad visual (la medición del constructo no tiene relación con el ajuste en el diseño y por tanto no se verá afectada).

Las pruebas que son diseñadas retomando el enfoque del diseño universal incluyen las siguientes características: “tienen poblaciones inclusivas a ser evaluadas, miden lo que pretenden medir, contienen un sesgo mínimo, contienen instrucciones y procedimientos claros y entendibles, tienen disposición al uso de acomodaciones,<sup>3</sup> son comprensibles y legibles” (Johnstone *et al.*, 2008: 25). La disposición al uso de acomodaciones significa que se “permitan cambios al formato sin que se cambie el significado o dificultad” (Johnstone, Altman y Thurlow, 2006). También se ha señalado la importancia del respeto a la diversidad de la población evaluada, que implica, por ejemplo, evitar distracciones<sup>4</sup> en las pruebas con contenido que pueda parecer descalificador o insultante a quienes las presentan (Zieky, 2006).

Lo anterior implica que los desarrolladores de las pruebas “tengan claridad sobre el constructo o constructos a ser medidos, incluyendo el objeto de la evaluación, el propósito con el que se usarán los puntajes, las inferencias que se harán sobre los puntajes, y las características de los sustentantes y subgrupos de la población objetivo de la prueba que pueden influenciar el acceso” (AERA *et al.*, 2014: 50).

El proceso que se sugiere que sigan las evaluaciones con este tipo de diseño comprende los pasos ilustrados en la figura 1.

El paso 1 requiere que, al contratar a externos para el desarrollo de instrumentos, se solicite que incluyan elementos de diseño universal. En este sentido, deben formular una propuesta que describa ampliamente la población objetivo, y que busque maximizar la cantidad de estudiantes de diversas características que puedan presentar las mismas pruebas. Esto deberá incluir a alumnos con algún tipo de discapacidad, con limitado dominio de la lengua en que se aplica la prueba, con necesidades educativas especiales, dentro de otros grupos de la población.

El paso 2 enfatiza la necesidad de que el constructo de la prueba sea claramente conceptualizado, incluyendo el número y tipo de ítems para cada una de las áreas. Las características de accesibilidad también deben ser abordadas por cada uno de los ítems, esto significa que

<sup>2</sup> En los países de habla inglesa se les denomina *English Language Learners*.

<sup>3</sup> Las acomodaciones son “cambios que se hacen en la administración de las pruebas para asegurar que los estudiantes con necesidades especiales puedan tener acceso al contenido de las pruebas – que puedan entender lo que se pregunta en los ítems o tareas de evaluación de tal manera que demuestren sus conocimientos” (Solano-Flores, 2012a: 5).

<sup>4</sup> Los distractores incluyen palabras o imágenes que puedan degradar, discriminar o ridiculizar a los grupos sociales a los que pertenecen o a personas que valoran (Zieky, 2006).

“respeten la diversidad de la población que será evaluada, siendo sensibles a sus características y experiencias (género, edad, etnicidad, estado socioeconómico, región, discapacidades, lenguaje)” (Johnstone *et al.*, 2006: 6). También requerirá: “evitar contenidos que den ventaja o desventajas injustas a estudiantes de algún subgrupo, y que minimicen los efectos de factores externos” (Johnstone *et al.*, 2006: 6).

**Figura 1** Proceso del Diseño Universal de las Evaluaciones



Fuente: elaboración propia basado en Johnstone *et al.*, 2006: 5-24.

El paso 3 requiere el involucramiento de equipos adicionales para revisar la evaluación. Estos equipos requieren la participación de miembros de diferentes grupos lingüísticos y con diversidad de discapacidades. Para esta revisión se deberá proveer a los equipos con las siguientes descripciones:

- Propósito de la prueba y estándares que mide cada ítem
- Descripción de los sustentantes (*v. gr.*, edad, región geográfica)
- Instrucciones de la prueba
- Formatos que se utilizarán en la prueba, incluyendo los formatos de respuesta
- Información sobre el uso de tecnología
- Políticas sobre modificaciones de acceso o acomodaciones (Johnstone *et al.*, 2006: 6-7)

El paso 4 incluye el uso de entrevistas cognitivas o de pensamiento en voz alta para analizar los ítems que los equipos hayan considerado que podrían ser problemáticos para los estudiantes.

Las entrevistas cognitivas son un método que permite “entender el pensamiento y enfoques de resolución de problemas utilizados por los estudiantes cuando responden los ítems de una prueba [...] permiten ver de qué manera el lenguaje, la presentación o las tareas asociadas con un ítem pueden distraer, confundir o desafiar adecuadamente a los estudiantes” (Johnstone *et al.*, 2008: 32).

El paso 5 sugiere una vuelta hacia los ítems para revisarlos utilizando la información obtenida en los pasos 3 y 4, y hacer modificaciones antes de la aplicación en campo.

El paso 6 comprende la puesta a prueba de los ítems en campo; se recomienda también poner a prueba las acomodaciones que hayan decidido hacerse para maximizar la cantidad de estudiantes de diversas características que las presenten.

El paso 7 involucra el análisis de los resultados de los ítems, utilizando las siguientes técnicas: clasificación de los ítems, correlación total de los ítems, funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés) utilizando tablas de contingencia, y DIF utilizando teoría de respuesta al ítem.

El paso 8 involucra la revisión final de la prueba, tomando en cuenta los resultados de los pasos anteriores. Como los otros momentos del proceso, se enriquece con el trabajo en equipo multidisciplinario.

El paso 9 comprende la aplicación definitiva de la prueba utilizando condiciones estándares y también las acomodaciones que se haya decidido implementar.

Con el paso 10 inicia de nuevo el proceso, se hacen análisis estadísticos, revisión de expertos, y entrevistas cognitivas sobre la próxima aplicación.

A lo largo de la descripción del proceso se señalaron los tres métodos principales que se utilizan en el diseño universal de las evaluaciones para revisar los ítems e identificar los que puedan ser problemáticos para poblaciones específicas: revisión por expertos, análisis estadísticos y entrevistas cognitivas. En el proceso se sugiere la utilización de los tres métodos. Algunas precisiones con respecto a cada uno de estos métodos se señalan en la continuación:

#### Revisión de expertos

Los paneles de revisión de expertos son utilizados regularmente en el proceso del diseño de las evaluaciones en gran escala, por lo general, con la intención de “minimizar el sesgo y promover la alineación de los ítems con los estándares curriculares” (Johnstone *et al.*, 2008: 28). Para disminuir los riesgos de una revisión no estructurada o comprensiones diferentes del sesgo del comité de revisión, investigadores especializados en diseño universal de las evaluaciones han generado un formato de revisión de los ítems, donde se precisan y describen los elementos a considerar, se provee un espacio para que los revisores señalen sus preocupaciones y sugerencias, y se da la opción para que sugieran una revisión posterior de los ítems por expertos o estudiantes con respecto al contenido, una discapacidad específica, lenguaje o cultura (Johnstone *et al.*, 2008).

En la tabla 3 se muestra el formato utilizado por Johnstone *et al.* (2008) en la revisión de expertos.

**Tabla 3** Formato para la revisión del diseño universal de los ítems

Considerations for Universally Designed Assessment Items							Subject ____	Grade ____	Test form ____	Item #s on this page_1-5____
Initials _____										
Star (*) areas of strenght and Check (✓) areas of concern for each item	Pass-age	Item #1	Item #2	Item #3	Item #4	Item #5	Describe Concerns and Suggestions for items and reading passages	Recommend review by expert or student in Content Area, Specific Disability, Language, Culture		
<b>Item measures its intended construct</b>										
<b>Item respects the diversity of the assessment population</b> <ul style="list-style-type: none"> <li>• Sensitive to test taker characteristics and experiences (gender, age, ethnicity, SES, region, disability, language)</li> <li>• Avoids content that might unfairly advantage or disadvantage any student subgroup</li> <li>• Other</li> </ul>								Expert review?	Student review?	
<b>Item has concise and readable text</b> <ul style="list-style-type: none"> <li>• Commonly used words (except vocabulary tested)</li> <li>• Minimum use of unnecessary words</li> <li>• Technical terms and abbreviations avoided unless tested</li> <li>• Sentence complexity appropriate for grade level</li> <li>• Questions to be answered identifiable</li> <li>• Other</li> </ul>										
<b>Item has a clear format to text</b> <ul style="list-style-type: none"> <li>• Standard typeface</li> <li>• Twelve (12) point minimum size for all print</li> <li>• High contrast between text and background</li> <li>• Sufficient blank space</li> <li>• Staggered right margins</li> <li>• Other</li> </ul>										
<b>Item has a clear visuals (use NA for none)</b> <ul style="list-style-type: none"> <li>• Visuals are needed to answer the question</li> <li>• Visuals have clearly defined features</li> <li>• High contrast between visuals and background</li> <li>• Visuals are clearly labeled</li> <li>• Other</li> </ul>										
<b>Item allows changes to format without changing meaning or difficulty (check allowed accommodations)</b> <ul style="list-style-type: none"> <li>• Braille or other tactile format</li> <li>• Sign language interpretation</li> <li>• Oral presentation</li> <li>• Assistive technology</li> <li>• Translation into other language</li> <li>• Other</li> </ul>										
Describe other considerations specific to item on back										

Fuente: Johnstone et al., 2008: 28.

## Análisis estadísticos

Se sugiere utilizar múltiples métodos estadísticos para valorar si los ítems son problemáticos para una población específica. Los métodos sugeridos son: clasificación de los ítems de acuerdo a su nivel de dificultad para diferentes grupos de población, correlación total de los ítems y pruebas de *t* para determinar la existencia de diferencias significativas entre poblaciones objetivo y de referencia con respecto a un ítem particular, funcionamiento diferencial de los ítems con tablas de contingencia, y funcionamiento diferencial de los ítems utilizando Teoría de Respuesta al Ítem (IRT, por sus siglas en inglés) (Johnstone *et al.*, 2008: 30-32).

El uso de múltiples métodos estadísticos, tiene que superar las limitaciones de cada uno de los métodos en particular. Por ejemplo, el análisis DIF, requiere grandes cantidades de alumnos para que sus resultados sean válidos, lo cual no permitiría que se realizaran análisis de grupos de alumnos con baja presencia en los exámenes. Si esto fuera así, seguramente se descartarían los análisis con esas poblaciones, o se dependería de un análisis simple para contar con más elementos para tomar una decisión.

## Entrevistas cognitivas

Las entrevistas cognitivas son un método de gran utilidad en la valoración del diseño universal de los ítems; permiten conocer los procesos cognitivos de los estudiantes al enfrentarse con los ítems; y complementan asimismo los análisis estadísticos, pues con éstos se identifican los ítems que resultan problemáticos para ciertos grupos de población, y a través de las entrevistas cognitivas se reconocen las razones (Johnstone *et al.*, 2008).

Se sugiere que después de realizar la entrevista cognitiva, se solicite información adicional a los estudiantes, especialmente para quienes el pensamiento en voz alta al enfrentarse y tratar de resolver un ítem haya resultado desafiante (Branch, 2001).

### 1.3. Diversidad cultural: un elemento esencial de la validez de las evaluaciones

La cultura de los estudiantes influencia la manera en que “construyen el conocimiento y significados a partir de las experiencias” (Solano-Flores y Nelson-Barber, 2001: 553). En cuanto a las evaluaciones de aprendizajes, la cultura influye en la forma en que los alumnos interpretan y responden a las actividades planteadas en las pruebas (Solano-Flores, 2011; Solano-Flores y Nelson-Barber, 2001). La falta de consideración de las características socioculturales de las poblaciones a las que están dirigidas las pruebas puede llevar a conclusiones imprecisas e inválidas sobre su desempeño. Lo anterior hace evidente la necesidad de incorporar en todo el proceso evaluativo “las influencias culturales que moldean la manera en la cual los estudiantes interpretan y resuelven problemas” (Solano-Flores y Nelson-Barber, 2001: 554).

La validez cultural de las evaluaciones ha sido definida como:

la efectividad con la que las evaluaciones [...] retoman las influencias socioculturales que moldean el pensamiento de los estudiantes y la manera en la cual comprenden los ítems [...] y les dan respuesta. Las influencias socioculturales comprenden los conjuntos de valores,

creencias, experiencias, patrones de comunicación, estilos de enseñanza y aprendizaje, y las epistemologías inherentes al contexto cultural de los estudiantes, y las condiciones socioeconómicas que prevalecen en sus grupos culturales (Solano-Flores y Nelson-Barber, 2001: 555)

Solano-Flores y Nelson-Barber (2001: 555) señalan cinco campos relacionados con la validez cultural que podrían contribuir a la mejora de las evaluaciones: "(a) epistemología de los estudiantes, (b) dominio de la lengua, (c) perspectivas culturales del mundo, (d) estilos culturales de comunicación y socialización, y (e) contexto de vida y valores de los alumnos".

La epistemología de los estudiantes se enfoca en la relación que establecen entre el contenido que se evalúa (o se enseña, en su caso) y sus propias experiencias; pues, por ejemplo, en las pruebas de ciencia se ha encontrado que "frecuentemente, las experiencias cotidianas parecen ser lo primero que viene a las mentes de los alumnos cuando responden los ítems" (Solano-Flores y Nelson-Barber, 2001: 559). El dominio de la lengua tiene que ver con las dificultades que enfrentan los alumnos que presentan una prueba en una segunda lengua y que pueden impedir que demuestren sus conocimientos; este punto será tratado con mayor detalle en el siguiente apartado. Las perspectivas culturales sobre el mundo, aluden a la relación que guardan los contenidos con las experiencias culturales de los alumnos, y requieren un entendimiento de la manera en la cual las culturas se acercan al conocimiento y sus propios conocimientos tradicionales. Los estilos de comunicación y socialización culturales, tienen que ver con la manera en la cual "se organiza el discurso en sus lenguas nativas" (Solano-Flores y Nelson-Barber, 2001: 562), que pueden influenciar el tipo de respuestas de los alumnos ante ciertos ítems, por ejemplo, su predisposición a expresar respuestas extensas o cortas. El contexto de vida y valores de los alumnos requiere la identificación de las "experiencias que son relevantes o representativas de su contexto cultural" (Solano-Flores y Nelson-Barber, 2001: 564).

Los campos señalados por Solano-Flores (2001) evidencian la necesidad de que en el proceso evaluativo se incorpore una etapa destinada a la "comprensión de la manera en que piensan los alumnos y las influencias socioculturales que moldean su pensamiento" (p. 567), y que esto se vincule con la elaboración y validación de los ítems. Esto tiene implicaciones en la conformación de los equipos que participan en las diferentes etapas del proceso evaluativo. Sería deseable, como se verá más adelante, que en los equipos que participan en el diseño de las pruebas haya docentes y otros integrantes de los grupos culturales a los cuales pertenecen los alumnos.

#### Diversidad lingüística:

##### los desafíos de la traducción de las pruebas

La inclusión de alumnos cuya primera lengua es distinta a aquella en que se aplican las pruebas tiene que ver con la atención a la diversidad lingüística en las evaluaciones (Solano-Flores, 2012). Se parte del reconocimiento que un dominio limitado de la lengua en la que se elaboran las pruebas "puede afectar negativamente el desempeño de los alumnos, y poner en riesgo la validez de las medidas de logro" (Solano-Flores, 2012a: 5).

Las acomodaciones que se implementan en las evaluaciones a este respecto, buscan garantizar la accesibilidad de las pruebas para estos grupos de población, de tal manera que los alumnos puedan demostrar sus conocimientos.



Un grupo de acomodaciones que pueden ser diseñadas para atender la diversidad lingüística tiene que ver con la traducción de las pruebas. La traducción es una tarea muy desafiante si se busca garantizar que “no cambien los constructos medidos, no se sugieran respuestas a los alumnos, y no den una ventaja a los alumnos que reciben las acomodaciones sobre quienes no son objeto de ellas” (Solano-Flores, 2012a: 5).

Los desafíos se presentan desde la identificación de la población objetivo pues difícilmente hay mediciones efectivas sobre el dominio de la lengua de los alumnos a ser evaluados que permitan reconocer los apoyos lingüísticos que necesitan. La confusión en la clasificación del nivel de dominio de la lengua en que se aplica la prueba, puede llevar a que las acomodaciones de traducción que se ofrezcan conlleven un perjuicio para los alumnos en vez de facilitar que demuestren sus talentos en la prueba.

Solano-Flores (2012a: 27), basado en Francis, Rivera, Lesaux, Kieffer y Rivera (2006), distingue los siguientes tipos de acomodaciones, de acuerdo con el apoyo lingüístico que ofrecen, ver la tabla 4.

**Tabla 4** Acomodaciones para alumnos cuya primera lengua no es el inglés  
(*English Language Learners*) de acuerdo con el apoyo lingüístico que ofrecen

Apoyo lingüístico indirecto	Apoyo lingüístico directo	
	En inglés	Acomodación de traducción
<ul style="list-style-type: none"> <li>• Ampliación del tiempo disponible para responder la prueba</li> <li>• Descansos entre sesiones de aplicación</li> </ul>	<ul style="list-style-type: none"> <li>• Glosarios en inglés</li> <li>• Diccionarios en inglés</li> <li>• Instrucciones leídas en inglés</li> <li>• Simplificación lingüística en inglés</li> <li>• Dictado de respuestas o uso de un escriba</li> </ul>	<ul style="list-style-type: none"> <li>• Versión de la prueba en la lengua nativa de los alumnos</li> <li>• Versión bilingüe de la prueba</li> <li>• Traducción de las instrucciones a la lengua nativa de los alumnos</li> <li>• Glosario bilingüe</li> <li>• Respuestas en la lengua nativa de los alumnos</li> <li>• Instrucciones leídas a los alumnos en su lengua nativa</li> </ul>

Fuente: Solano-Flores, 2012a: 27.

Los métodos y procedimientos utilizados en las traducciones y su revisión deben considerar “las características de los ítems incluidos en las pruebas y las características de la población objetivo” (Solano-Flores, 2012a: 37). Estos métodos pueden variar de acuerdo a la acomodación a emplear. Solano-Flores (2012a: 44-47) ha desarrollado modelos de traducción para las siguientes acomodaciones: versión de la prueba en la lengua nativa de los alumnos, versión bilingüe de la prueba, traducción de las instrucciones a la lengua nativa de los alumnos, y glosario bilingüe.

Se precisa que las decisiones sobre las acomodaciones a emplear sean precedidas por una caracterización de la heterogeneidad lingüística de la población, la cual lleve a determinar qué grupos pueden ser sujetos a una traducción de la prueba. Los criterios utilizados en esa valoración son: frecuencia (porcentaje de usuarios de lengua nativa por estado y a nivel nacional), factibilidad (que existan traductores altamente calificados) y estabilidad geográfica de la población que habla la lengua nativa (tiene que ver con la movilidad de los grupos poblacionales, ya que la traducción de las pruebas destinadas a grupos poblacionales que migren constantemente implicará dificultades para su aplicación y obtención de resultados esperados con respecto a la medición del desempeño) (Solano-Flores, 2012a). Además se deberá valorar

el tipo de acomodación de acuerdo a la probabilidad de cubrir las siguientes dimensiones de validez y justicia de las evaluaciones (Solano-Flores, 2012a: 28-32):

- Seguridad de los sustentantes que no requieran la acomodación pero que la reciban en caso de una clasificación inadecuada.
- Sensibilidad a las necesidades lingüísticas individuales de los sustentantes. La posibilidad de que los sustentantes puedan utilizar las acomodaciones de acuerdo a sus necesidades individuales.
- Fidelidad de la implementación. Tiene que ver con la capacidad de que la acomodación pueda ser implementada de forma estandarizada entre los sustentantes.
- Facilidad de uso. Está relacionada con la demanda cognitiva adicional que pueden representar las acomodaciones de traducción para los sustentantes.

Aunque cada acomodación de traducción requiere procedimientos específicos, comparten los procesos que se muestran en tabla 5.

**Tabla 5** Procesos comunes en las acomodaciones de traducción

---

**Actividades de preparación de la traducción**

Comprende el establecimiento de relaciones con las agencias involucradas en el desarrollo de la prueba para promover un entendimiento de la complejidad de la traducción y la definición del plan de trabajo. En este proceso participan equipos multidisciplinarios.

---

**Traducciones independientes y conciliación de traducciones**

Es común que la traducción inicial la hagan varios traductores independientes, y un traductor adicional haga la conciliación de las traducciones, resolviendo las diferencias y garantizando que se mantenga el significado y nivel de dificultad.

---

**Revisión de la traducción**

Un equipo multidisciplinario se hace responsable de la creación de las versiones finales de las traducciones. En conjunto revisan los ítems traducidos, considerando dimensiones del diseño del reactivo, del lenguaje y del contenido a evaluar, para decidir las mejoras a implementar.

---

**Entrevistas cognitivas**

Se implementan para una muestra de ítems; los alumnos con limitado uso de la lengua en la cual se aplica la prueba participan en el proceso. Estas entrevistas proveen información que ayuda en la mejora de las traducciones y también para determinar el tiempo adicional que requerirán los alumnos para dar respuesta a la prueba.

---

**Versión final de la traducción**

Esta versión debe ser acompañada de indicaciones sobre su formato y producción.

---

**Actividades de seguimiento**

Involucra acciones encaminadas a asegurar que la prueba impresa tenga las características especificadas.

---

Fuente: adaptado de Solano-Flores, 2012a: 42-54.<sup>5</sup>

Los equipos multidisciplinarios que participan en diferentes etapas del proceso de generación de las pruebas traducidas comprenden, por ejemplo, traductores, docentes bilingües, maestros de la disciplina que será objeto de evaluación, especialistas en el contenido a evaluar, desarrolladores de pruebas (psicómetras) y sociolingüistas (Solano-Flores, 2012a: 39-40).

Para disponer de reactivos accesibles a las variaciones del lenguaje, el marco conceptual que utilizan en el *Smarter Balanced* está apoyado las perspectivas teóricas señaladas en la tabla 6 (Solano-Flores, Shade y Chrzanowski, 2014):

<sup>5</sup> Ver Solano-Flores (2012a: 58) para mayor información sobre las dimensiones del diseño de los ítems, lenguaje y contenido, que deben ser incluidas en la revisión de los ítems traducidos.

**Tabla 6** Perspectivas teóricas en las que se apoya el *Smarter Balanced*

<b>Teoría sociocultural</b>
<p>Teoría que examina el pensamiento como un fenómeno cultural. Permite la identificación de aspectos sociales y culturales que influyen cómo los estudiantes interpretan los ítems. Esta teoría está fundamentada en la premisa de que los individuos no actúan directamente sobre el mundo, sino que la mente humana está mediada por herramientas (simbólicas y concretas) y actividades.</p> <p>La perspectiva sociocultural provee un lente para examinar cómo la cultura y los niveles de participación social conforman el significado que los estudiantes le dan a los reactivos. Como producto de la cultura, las pruebas asumen que los estudiantes tienen ciertos niveles de participación social en determinados ámbitos culturales. Diferentes conjuntos de experiencias culturales influyen cómo los estudiantes acceden a los reactivos.</p>
<b>Sociolingüística</b>
<p>Que consiste en el estudio de la lengua como un reflejo de la elección y la variación social. Examina la relación entre las variables lingüísticas y categorías sociales (por ejemplo la raza, el sexo y la clase), y cómo los individuos co-construyen sus identidades a través del uso de estas variables durante la interacción social.</p> <p>El pensamiento actual en el campo ya no trata a las lenguas, las variedades del lenguaje y el lenguaje de las comunidades como algo delimitado y uniforme; los dialectos de una lengua varían según las características sociales de la comunidad lingüística, tales como la clase social, la ubicación geográfica y la edad. Es decir, todo el mundo habla un dialecto.</p> <p>Debido a las diferencias en sus experiencias culturales, cada alumno tiene un conjunto único de fortalezas y debilidades en el inglés —como segunda lengua— y en su lengua materna. Como consecuencia, la eficacia de los recursos para la accesibilidad de los reactivos es, en gran medida, en función de qué tan probable es que funcione según lo previsto para la mayoría de los estudiantes hacia los que están dirigidos.</p>
<b>Semiótica social</b>
<p>Estudia cómo el significado se transmite y construye a través de diversas formas de representación de la información, y las formas en que este proceso está mediado por la cultura. Permite examinar las características de los reactivos y de los recursos de accesibilidad de los ítems como formas alternativas de representación de la información.</p> <p>Para interpretar correctamente los reactivos y los recursos de accesibilidad de los ítems, los examinados deben ser capaces de procesar e integrar los diferentes sistemas de significado. Para hacer frente a la noción de que múltiples modos semióticos trabajan juntos para comunicar significado, los desarrolladores de prueba deben tener un enfoque integral en el diseño de reactivos y de sus recursos de accesibilidad. Este enfoque comprensivo toma en cuenta las múltiples características de los reactivos elementos y su interacción. Además, los reactivos de una prueba deben ser piloteados con muestras representativas de diferentes segmentos de la población objetivo para asegurar que los estudiantes de diferentes grupos lingüísticos y culturales interpretan los reactivos y sus recursos de accesibilidad de forma similar, según lo previsto por los desarrolladores de la prueba.</p>
<b>El lenguaje como error en la medición</b>
<p>La perspectiva del lenguaje como una fuente de error en la medición establece un vínculo entre la variación del lenguaje y la varianza del error en las pruebas. Combina la visión del lenguaje como un proceso de comunicación mediado por el contexto y los métodos de la teoría de la Generalizabilidad (teoría psicométrica del error de medición).</p> <p>La evaluación puede ser vista como un proceso de comunicación entre el estudiante y el sistema de evaluación. Debido a que la evaluación se realiza a través del lenguaje, toda evaluación es, hasta cierto punto, una medida de la competencia en el lenguaje. En consecuencia, el error de medición del logro académico es, en gran medida, debido a la variación lingüística. Las facetas de la variación lingüística incluyen el dominio del idioma y el dialecto; así, desde esta perspectiva, los recursos de accesibilidad a los reactivos intentan minimizar el error de medición debido a esas fuentes de variación.</p>
<b>Teoría del error de traducción de pruebas</b>
<p>Examina cómo la traducción de una prueba se conforma principalmente por el hecho de que las lenguas codifican la experiencia de manera distinta. Esta teoría contribuye a la noción del efecto agregado del error de traducción del test. Múltiples errores insignificantes de traducción, que por separado no afectarían el rendimiento del estudiante, pueden tener un efecto negativo sustancial cuando coinciden en un reactivo. Este conocimiento pone de relieve la importancia de examinar las características de los recursos de accesibilidad de los reactivos de manera holística.</p>

Fuente: traducción propia y adaptación de Solano-Flores, Shade, *et al.*, 2014.

La naturaleza multidimensional del error de traducción apunta a la necesidad de profesionales de múltiples disciplinas (por ejemplo, traductores, lingüistas, maestros y expertos en el contenido y la evaluación) para participar en el proceso de traducción de una prueba y su revisión. Este enfoque contrasta con el enfoque tradicional de la traducción de pruebas, que se basa en la experiencia de un número limitado de traductores.

Para incrementar la accesibilidad de los reactivos para estudiantes cuya primera lengua es distinta a aquella en la que se aplica la prueba, el *Smarter Balanced* ha definido los principios que se describen en la tabla 7 (Solano-Flores, Shade, *et al.*, 2014):

**Tabla 7** Principios para incrementar la accesibilidad de los reactivos del Smarter Balanced

<b>Consistencia</b>
Condición en la que los elementos con funciones similares y significados comparten también una apariencia similar de los reactivos en la prueba. También se refiere a la condición en la cual la apariencia de esos elementos coincide con la experiencia previa de los estudiantes.
<b>Simetría</b>
Es la condición en la cual las entidades con la misma importancia reciben el mismo nivel de consideración y énfasis. Cuando las lenguas son tratadas de manera asimétrica, puede afectar la eficacia de la accesibilidad de los reactivos y, en última instancia, la validez de la prueba.
<b>Contraste</b>
Es la condición en la cual los elementos con diferentes funciones son representados de manera distinta. Ayuda a los usuarios a advertir diferencias importantes entre los elementos de la prueba.
<b>Significancia</b>
Es la condición en la que los estudiantes tienden a relacionar el contenido de los reactivos de la prueba con la información contextual que han obtenido de su experiencia personal.
<b>Restricciones de diseño</b>
Son límites impuestos a ciertos parámetros que no pueden excederse en el diseño de los reactivos o sus traducciones. Estas restricciones están destinadas a minimizar la variación innecesaria en las características lingüísticas de los reactivos. Incluyen: longitud del texto, características gramaticales, uso restringido de constituyentes y nombres ficticios de los personajes.
<b>Adaptación y diseño desde el principio</b>
La adaptación implica la modificación de algunas características de un texto para cumplir requisitos específicos. El diseño desde el principio se hace cuando la adaptación no es posible, e implica la creación del texto nuevamente.
<b>Conveniencia pragmática</b>
Es la condición en la que las características de los reactivos son apropiadas en el contexto, lo que asegura que los estudiantes los interpretan según lo previsto por sus desarrolladores.
<b>Estandarización</b>
Es el conjunto de acciones adoptadas con el fin de asegurar que todos los reactivos del mismo tipo tienen el mismo formato y se desarrollan con el mismo procedimiento.

Fuente: traducción propia y adaptación de Solano-Flores, Shade, *et al.*, 2014.

## 2. RELACIÓN DE LA ACCESIBILIDAD DE LAS PRUEBAS CON LAS OPORTUNIDADES DE APRENDIZAJE DE LOS ALUMNOS Y EL CURRÍCULUM

En torno al desarrollo alcanzado en la promoción de la justicia de las evaluaciones, algunos autores postulan que ésta se relaciona con la justicia en cuanto a las oportunidades de acceso, y también con los planteamientos curriculares (Stobart, 2005). La manera en la que los estudiantes entiendan las consignas incluidas en las pruebas y les den respuesta se relaciona con las

oportunidades de aprendizaje que hayan tenido dentro de la escuela y con el vínculo entre el currículo y su contexto sociocultural.

En la tabla 8, retomada de Gipps y Stobart (2010), se muestran preguntas relacionadas con la justicia y equidad de las evaluaciones, las cuales dejan ver la relación que guardan las características de acceso y el currículo con la evaluación de los aprendizajes.

**Tabla 8** Preguntas relacionadas con la justicia y equidad de las evaluaciones

Preguntas de acceso a oportunidades educativas	Preguntas curriculares	Preguntas de evaluación
<ul style="list-style-type: none"> <li>• ¿Qué es enseñado y por quién?</li> <li>• ¿Hay diferencias en los recursos disponibles para los diferentes grupos?</li> <li>• ¿Qué se incorpora en la enseñanza sobre las culturas de los alumnos?</li> </ul>	<ul style="list-style-type: none"> <li>• ¿Qué conocimiento se enseña?</li> <li>• ¿Por qué se enseñan de una manera particular a un grupo específico?</li> <li>• ¿De qué manera se permite que las historias y culturas de la gente de color y las mujeres sean enseñadas de manera responsable?</li> </ul>	<ul style="list-style-type: none"> <li>• ¿Qué conocimientos se evalúan y a qué logro corresponden?</li> <li>• ¿La forma, contenido y modo de evaluación son apropiados para los diferentes grupos e individuos?</li> <li>• ¿El conocimiento cultural se refleja en las definiciones de desempeño?</li> <li>• ¿De qué manera el conocimiento cultural media las respuestas de los individuos a las evaluaciones de manera tal que pueda alterar el constructo medido?</li> </ul>

Fuente: Gipps y Stobart, 2010: 58.

A excepción de la primera pregunta de la columna relativa a la evaluación, el resto están comprendidas dentro de los aspectos a considerar en la validez cultural de las evaluaciones. Las preguntas de acceso a oportunidades educativas y las curriculares, aluden a aspectos que también influyen en la manera en la cual los alumnos dan respuesta a las evaluaciones.

En cuanto al acceso a oportunidades educativas, Gee (2003) sostiene que: “[s]i dos niños son evaluados con respecto a algo en lo que no han tenido las mismas oportunidades para aprender, la evaluación es injusta (a menos que el propósito de la evaluación sea mostrar las disparidades en las oportunidades de aprendizaje)” (p. 28). En los estándares de AERA *et al.* (2014) se reconoce también que las diferentes oportunidades educativas pueden tener implicaciones con respecto a “la manera en que se interpretan los puntajes y los usos que se pretende hacer de ellos” (AERA *et al.*, 2014: 56); enfatizan que “la falta de consideración de las oportunidades de aprendizaje puede llevar a un mal diagnóstico de los alumnos, asignación o ubicación inadecuada a ciertos servicios, que podrían tener consecuencias significativas para los individuos” (AERA *et al.*, 2014: 57). Se precisa que las oportunidades de aprendizaje se convierten en un asunto de justicia cuando las autoridades educativas proveen diferentes oportunidades a ciertos individuos y después hacen responsables a estos individuos por su desempeño, esto ocurre particularmente con las pruebas de alto impacto, de cuyos resultados puede depender el ingreso de los alumnos a un nivel educativo superior. Ante esto, se sugiere que antes de que se implementen consecuencias negativas por los resultados de una evaluación, “se asegure que se cuenta con evidencias de que a los alumnos se les proveyó tanto un currículum como enseñanza que comprendiera los constructos medidos en la prueba” (AERA *et al.*, 2014: 57).

Una postura determinante como la de Gee (2003) podría llevar a conclusiones sobre la imposibilidad de generar evaluaciones justas o más justas, ya que lograr una equivalencia total de las oportunidades educativas entre los diversos grupos de estudiantes que se desee evaluar

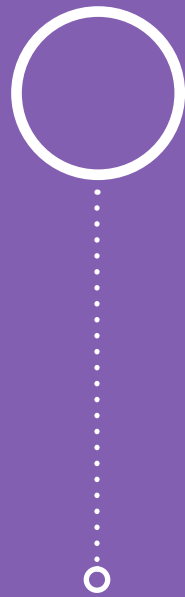
pareciera, igualmente imposible de lograr. Ante esto, es importante recordar, por un lado, la dificultad que se ha encontrado para definir oportunidades de aprendizaje, igualmente para su medición, pues, por ejemplo, aunque el currículo indique que ciertos contenidos que están incluidos en la prueba se traten en las aulas, una medición precisa de la cobertura de este contenido para cada estudiante sería prácticamente imposible (AERA *et al.*, 2014).

Algunos elementos que podrían contribuir a tener evaluaciones más justas, con respecto a las oportunidades de aprendizaje son:

- Incluir acercamientos a las oportunidades de aprendizaje de los diferentes grupos culturales y lingüísticos evaluados en las aplicaciones de logro educativo, de tal manera que permitan tener un mejor entendimiento de los contextos de aprendizaje, y retroalimenten las decisiones en materia de política educativa en cuanto a prácticas docentes y escolares, recursos educativos, y otras condiciones para la enseñanza y el aprendizaje.
- Proveer mayor información a las escuelas y los sustentantes para comprender los propósitos de las evaluaciones, su estructura y las características de la aplicación. De acuerdo con Stobart (2005), esto permitiría que estuvieran mejor preparados para presentar las pruebas. Algunas estrategias que sugiere son: proveer orientaciones sobre la prueba; entregar reportes con datos estadísticos que den retroalimentación sobre los resultados obtenidos; mantener el estilo y formato de la prueba; devolver a las escuelas los cuadernillos.

En otros sistemas de evaluación se ha incorporado un proceso en el que previo a la aplicación de la prueba, se realiza una supervisión sobre las oportunidades de aprendizaje en las escuelas, y la aprobación de los candidatos a presentar el examen depende de los resultados que se obtengan en ella (Stobart, 2005).

Las preguntas curriculares plantean un desafío en cuanto a la articulación entre el currículo y las culturas de los grupos poblacionales a los que se dirigen. Esta articulación es indispensable en la búsqueda de que la oferta educativa sea relevante para los alumnos. Con relación a la evaluación y el currículo, Stobart (2005) sugiere que al definir los estándares y los contenidos de las pruebas, los equipos a cargo se pregunten si realmente los conocimientos que se evalúan son necesarios para los alumnos y que se analicen las razones por las cuales se mantienen ciertos contenidos en las pruebas.



# Prácticas internacionales de atención a la diversidad

## Países

### 1.1. DESCRIPCIÓN GENERAL

El sistema educativo chileno (SEC) se divide en cuatro niveles educativos: enseñanza parvularia, educación básica, educación media y educación superior. En Chile, existen 16 474 establecimientos proveedores de educación que atienden a casi cinco millones de alumnos matriculados, 738 656 niños en el nivel parvulario, 2 001 942 en educación básica, 1 042 395 en la educación media y 1 184 805 en el nivel superior. Para dar atención a la demanda educativa preuniversitaria, Chile cuenta con 171 882 docentes de aula que, junto con el personal directivo, de apoyo pedagógico y de la mesoestructura, integran un total de 207 671 docentes (Centro de Estudios MINEDUC, 2013; MINEDUC, 2013).

En 1988 se creó el Sistema de Medición de la Calidad de la Educación (SIMCE), a partir del cual inicia el desarrollo de evaluaciones externas de escuelas en el SEC. El SIMCE nació con el propósito de proveer de información relevante al SEC y sus principales actores. En sus inicios el SIMCE evaluó el rendimiento escolar con el propósito de monitorear el sistema educativo, valorar la calidad de la educación ofrecida y orientar las actividades de formación docente (Agencia de Calidad de la Educación, 2013b).

A partir de 1992 y hasta la primera mitad de 2012 el Ministerio de Educación (MINEDUC) asumió la organización e implementación del SIMCE con nuevas directrices. Pronto, inició la medición del logro académico de estudiantes que cursaban 4° y 8° básico en las asignaturas de Matemáticas y Español de forma censal, y en Ciencias Naturales e Historia de forma muestral. En 1994 se incorporó la medición de estimadores psicosociales, tales como la aceptación de la labor educacional, el desarrollo personal, las actitudes hacia el ambiente y el desarrollo de la creatividad.

Desde 1995 el SIMCE publica en los medios de comunicación los resultados de sus pruebas y otras variables, con el objetivo de proveer información a los padres de familia para la toma de decisiones sobre la educación de sus hijos, lo que atrajo la atención no sólo de docentes y directivos, sino de la opinión pública en general.

En 2003, con la conformación de la Comisión para el Desarrollo y Uso del Sistema de Medición de la Calidad de la Educación (Comisión SIMCE), cuyo propósito fue tener un diagnóstico sobre el sistema y la información que proporciona, así como entregar propuestas de mejoramiento orientadas a perfeccionar su impacto en la calidad de los aprendizajes, se reforzó la idea de que el SIMCE es un sistema metodológicamente sólido que ha permitido centrar la atención en los resultados de aprendizaje (Unidad de Currículum y Evaluación, 2007, 2008, 2009, 2010, 2011b, 2012g). Los informes de la Comisión SIMCE señalaban que el uso de los resultados de la evaluación no siempre ha sido coincidente con los propósitos para los que fueron diseñados los instrumentos de medición, debido principalmente a que la información que se proporciona a los docentes no permite orientar las prácticas pedagógicas. Derivado de ello, la Comisión



recomendó enfocar la medición en los logros de aprendizaje. Asimismo, recomendó recoger y comunicar los resultados, considerando variables de la población, como son la situación socioeconómica, el género, la ubicación geográfica, etc. (Agencia de Calidad de la Educación, 2013b).

A partir del segundo semestre de 2012 el SIMCE forma parte de la Agencia de Calidad de la Educación, la cual constituye un servicio público que funciona de manera descentralizada del MINEDUC, tiene consejo exclusivo, está dotada de personalidad jurídica, patrimonio propio y se relaciona con el Presidente de la República a través del MINEDUC.

A pesar que en Chile existen evaluaciones de otros actores, en particular de los docentes del SEC del nivel básico, tales como el Sistema de Evaluación del Desempeño Profesional Docente, con evaluaciones obligatorias cada cuatro años para verificar el cumplimiento de los estándares del Marco de la Buena Enseñanza, o la Asignación de Excelencia Pedagógica, un programa voluntario para reconocer el mérito profesional de maestros de primaria y secundaria que se someten a una evaluación basada en la revisión de un portafolio con planes de clases, grabación en video de una lección, y una prueba de conocimientos disciplinarios y pedagógicos (Manzi, González y Sun, 2011), no se retoman en este informe porque no se identificaron elementos de atención a la diversidad que resultaran especialmente sobresalientes, y más bien, se recuperan las evaluaciones de logro educativo donde sí se advierten esfuerzos importantes por considerar la diversidad de los evaluados, aunque sólo algunos tipos de diversidad.

## 1.2. Evaluaciones de logro del SIMCE

La Agencia de Calidad de la Educación (ACE) recolecta información a través de pruebas cuyo propósito principal es evaluar el aprendizaje de los estudiantes en los contenidos y habilidades del currículo nacional en diferentes asignaturas o áreas y niveles para contribuir al mejoramiento de la calidad y equidad de la educación (Agencia de Calidad de la Educación, 2013b). Además de las pruebas referidas al currículo, el proceso también recoge información sobre docentes, estudiantes, padres y apoderados a través de Cuestionarios SIMCE (Unidad de Currículum y Evaluación, 2012b, 2012c, 2012d, 2012e, 2012f). Esta información se utiliza para relacionar los resultados de los estudiantes en las pruebas SIMCE con el contexto escolar y social en que éstos aprenden, y para evaluar el grado de cumplimiento de cuatro de los ocho indicadores que forman parte de los Otros Indicadores de Calidad Educativa (OICE) (Unidad de Currículum y Evaluación, 2013, 2014b). Los cuatro OICE que se construyen a partir de los cuestionarios SIMCE son: Autoestima Académica y Motivación Escolar, Clima de Convivencia Escolar, Participación y Formación Ciudadana, y Hábitos de Vida Saludable. Los restantes OICE, construidos mediante registros del ministerio de educación chileno son: Asistencia Escolar, Retención Escolar, Equidad de Género, y Titulación Técnico-Profesional (Unidad de Currículum y Evaluación, 2014b).

En el nivel educativo básico las asignaturas que actualmente evalúa SIMCE son: Lenguaje y Comunicación (Comprensión de Lectura y Escritura), Matemática, Ciencias Naturales, Historia, Geografía y Ciencias Sociales. En este nivel las pruebas SIMCE se aplican a estudiantes de 2º, 4º, 6º y 8º grados. Las pruebas de Lengua y Matemáticas se aplican todos los años y en todos los grados; el resto de las asignaturas se distribuyen alternadamente en algunos años y grados educativos, de acuerdo al Calendario de Evaluaciones de la ACE (Agencia de Calidad de la Educación, 2014).

Otra característica de las evaluaciones SIMCE es que verifican el grado de cumplimiento de los objetivos generales a través de la medición de *estándares de aprendizaje* (Unidad de Currículum y Evaluación, 2014a). Es así que los resultados de las pruebas son reportados en cuanto al cumplimiento de esos estándares, los que son elaborados por el MINEDUC, y reflejan una descripción de lo que los estudiantes deben saber y poder hacer para demostrar el cumplimiento de los objetivos de aprendizaje estipulados en el currículum vigente.

Las pruebas SIMCE son matriciales, es decir, ante la imposibilidad práctica de que un estudiante responda reactivos que refieran a todo el currículum, se generan diferentes formas o cuadernillos de la prueba que contienen sólo una muestra de reactivos, lo que impide ofrecer resultados de manera individual; no obstante, las pruebas están diseñadas para que dentro de un mismo centro escolar se tenga una cobertura suficiente de las temáticas del currículum de tal manera que sea posible devolver resultados por escuela.

### 1.3. Prácticas destacables de atención a la diversidad

Dentro de la gama de evaluaciones que realiza la ACE, existen dos prácticas de atención a la diversidad que es relevante destacar; por un lado, la consideración del contexto socioeconómico, el género y la ruralidad de los estudiantes y, por otro, la adaptación de las pruebas para alumnos con discapacidad sensorial (DS). En los siguientes apartados se desarrolla cada una de estas prácticas.

#### 1.3.1. Consideración del contexto socioeconómico, el género y la ruralidad

En las pruebas SIMCE regulares se atiende la diversidad en varias de las etapas del proceso de diseño de los instrumentos, análisis de resultados y su difusión. Las subpoblaciones consideradas son principalmente las que derivan de las características socioeconómicas de las familias de los estudiantes y su género; también se presta atención a la diversidad de contextos en función del tamaño de las localidades donde habitan los alumnos, aunque con menor intensidad. Enseguida se abordan algunas de las etapas que muestran la atención prestada a la diversidad en el marco de esta evaluación.

#### Diseño de los instrumentos de evaluación

En el desarrollo de las pruebas SIMCE regulares, una vez que se definen los dominios a evaluar y se elaboran las tablas de especificaciones y las primeras versiones de los reactivos, la relevancia de los tópicos del currículum que se abordan y la suficiencia de su cobertura son sometidas a jeece. Además:

se elaboran criterios básicos para la selección y/o generación de estímulos y contextos que forman parte de las pruebas SIMCE, para todas las áreas y niveles evaluados. Estos criterios aluden a la consideración de que todos los estudiantes que rinden las pruebas tengan acceso igualitario a la comprensión de los estímulos y contextos, independientemente de su género, raza, religión, condición social, económica, cultural, física u otra, y que se tenga especial cuidado con que los estímulos y contextos no sean emocionalmente perturbadores, evadiendo temas o situaciones traumáticas o violentas,

o temas controversiales. Se considera además, que los estímulos y contextos deben ser adecuados al nivel escolar de los estudiantes que los enfrentarán. Esto implica que deben evitarse tanto estímulos y contextos complejos o distantes en términos temáticos, lingüísticos o de otra índole, así como los extremadamente simples o que pueden abordarse mediante conocimiento que proviene fuera de la escuela, alineados a los contenidos curriculares de cada nivel. Por último, se busca que los diferentes sectores sociales o culturales que constituyen la sociedad chilena se vean representados en la prueba, y evitar, por tanto, que ésta sea el reflejo de uno solo de estos sectores en particular. De esta manera, los estímulos y contextos que se seleccionen o generen deben abarcar, en su conjunto y en la medida de las posibilidades, la diversidad geográfica, cultural y social de nuestro país, así como representar a mujeres y hombres en sus distintas etapas etarias y en distintas situaciones (familiares, laborales, escolares, etc.) (Agencia de Calidad de la Educación, 2013b: 25).

En los comités de validación de reactivos se incluye una variedad de figuras relacionadas directamente con la docencia, como especialistas en didáctica, en evaluación en la disciplina y personal propio de la ACE especialista en evaluación educativa, de tal manera que se disponga de una visión multidisciplinaria que permita identificar aquellos reactivos que podrían incluir elementos que dificulten su resolución por ciertos grupos poblacionales.

Una vez que los reactivos pasan por el comité de validación y después de hacer los ajustes necesarios para aquellos en los que se identificaron posibles fuentes de sesgo cultural, se pilotea su funcionamiento con una muestra reducida de estudiantes. A partir de los resultados, los reactivos se valoran cualitativamente mediante un panel de expertos. Los que superan esta revisión son susceptibles de incluirse en un segundo piloteo realizado con una muestra representativa de estudiantes del país.

Al realizar el piloteo con una muestra representativa del ámbito nacional aumenta la posibilidad de que en forma previa a que un reactivo se incluya en las pruebas censales, se identifique la presencia de algún sesgo cultural. Ésta es una acción clara de atención a la diversidad, porque al incluir estudiantes con las diversas características que interesa considerar (urbano/rural, grupo socioeconómico [bajo, medio-bajo, medio, medio-alto, alto], lengua, hombre/mujer, entre otros rasgos), se podrá disponer de información mucho más precisa para seleccionar sólo aquellos ítems que no presenten algún sesgo hacia alguna de las subpoblaciones chilenas.

Después de pilotear los reactivos en campo se realizan múltiples análisis psicométricos tanto de la Teoría Clásica de las pruebas como de la Teoría de Respuesta al Ítem (TRI), con el propósito de identificar reactivos mal calibrados o que, en los hechos, no formen parte del dominio que se pretende evaluar. Al final de esta etapa se clasifican los ítems en tres categorías según la acción que se deba tomar: aceptar, modificar y rechazar.

Además de los análisis psicométricos, una estrategia fundamental para atender la diversidad es la realización de análisis para valorar el *funcionamiento diferencial de los reactivos* (DIF, por sus siglas en inglés). Como señala la ACE (2013b: 30): “El DIF ocurre cuando estudiantes con un mismo nivel de habilidad estimado tienen una probabilidad distinta de responder correctamente al ítem sólo por el hecho de pertenecer a un grupo determinado. Por ejemplo, un ítem cualquiera presentará DIF de género si los varones tienen mayor o menor probabilidad de responder correctamente que las mujeres, aun teniendo el mismo nivel de habilidad en la disciplina evaluada”.

Estos análisis para identificar el DIF son indispensables pero insuficientes para identificar reactivos con sesgo cultural. Para identificarlo a cabalidad, el análisis DIF debe ser secundado por análisis cualitativos para valorar la estructura y contenido de los reactivos, así como los procesos de respuesta que emplean los evaluados. Para realizar dicha valoración cualitativa en las pruebas SIMCE se hace lo siguiente:

- Si un ítem tiene una dificultad alta desde la perspectiva psicométrica, se revisa que esté redactado en forma comprensible para todos los estudiantes del nivel evaluado. Para esto, son útiles las revisiones de los registros de observación del proceso de aplicación, ya que permiten saber si hubo falta de comprensión, ya sea de la totalidad del ítem, parte del enunciado o su clave. Además, se revisa que los enunciados no induzcan al error engañosamente, que el enunciado incluya toda la información necesaria para responder el ítem correctamente, y que la clave sea completamente correcta y sus distractores completamente incorrectos.
- Si se observa en la curva un distractor que es seleccionado sólo por alumnos con habilidad “media” o sólo por alumnos con habilidad “media-alta”, se considera la posibilidad de que el distractor corresponda a una respuesta parcialmente correcta. Los distractores con este comportamiento son revisados para descartar que exista algún grado de corrección en su contenido, o una forma de interpretarlos como correctos.
- Si un ítem es muy fácil desde la perspectiva métrica, se revisa que los distractores sean verosímiles y plausibles, en otras palabras, que resulten razonables y atractivos para los alumnos que no han logrado el objetivo de aprendizaje evaluado por el ítem. En el análisis de estos ítems, también se verifica que en él no se induzca a escoger la clave, esto es, que no se den “pistas” para responder correctamente (Agencia de Calidad de la Educación, 2013b: 30).

Por otro lado, además de las pruebas, se diseñan cuestionarios de contexto que son aplicados a docentes, padres de familia y los mismos estudiantes (Unidad de Currículum y Evaluación, 2012b, 2012c, 2012d, 2012e, 2012f). Los cuestionarios SIMCE tienen dos grandes propósitos: 1) recoger información de contexto que permita analizar los resultados educativos obtenidos por los estudiantes y los establecimientos, en función de variables internas y externas a estos últimos, que podrían resultar relevantes, y 2) recabar información que permita medir y evaluar a los establecimientos en los OICE, que son complementarios a los resultados de aprendizaje (Unidad de Currículum y Evaluación, 2014b).

La selección de las variables que se incluyen en los cuestionarios SIMCE obedece a cinco criterios sustantivos y cuatro metodológicos. Los primeros consisten en:

- Alineación con objetivos de la Agencia de Calidad de la Educación.
- Aporte a la comprensión de resultados de aprendizaje y de OICE fundada en la evidencia.
- Aporte a la contextualización de los resultados de aprendizaje y de los OICE, tal como el nivel socioeconómico de las familias de los estudiantes, el género de los evaluados y el tipo de localidad donde viven.
- Útil al diseño de políticas y a la toma de decisiones de los establecimientos.
- Factible de ser modificado, en caso de ser necesario.
- Aceptación pública y no atentar contra la privacidad ni la dignidad de las personas.

Los cuatro criterios metodológicos consisten en:

- Factibilidad de obtención de la información mediante cuestionarios autoaplicados.
- Cumplimiento de estándares profesionales de confiabilidad y validez.
- Variabilidad y relevancia de la información de contexto de un periodo a otro.
- Extensión de los cuestionarios y restricción presupuestaria de la Agencia.
- Como puede apreciarse, nuevamente el establecimiento de criterios para la inclusión de variables que ayuden a situar los resultados de aprendizaje constituye una acción que aporta a la atención de la diversidad en las evaluaciones de gran escala.

### Aplicación de los instrumentos

Las pruebas de comprensión lectora y matemáticas se aplican de forma censal a los alumnos que acuden a los centros educativos *regulares*, con excepción de los alumnos diagnosticados con necesidades educativas especiales (NEE) permanentes, para los cuales se han diseñado pruebas especiales y que se reportan en un apartado específico de este informe.

Las pruebas de escritura y educación física se aplican de forma muestral, para lo cual se utiliza un diseño por estratos que considera la modalidad educativa, el nivel de logro de los centros escolares en las otras pruebas SIMCE, y la ruralidad del establecimiento escolar.

El proceso de aplicación conlleva una serie de etapas que comprenden desde el reclutamiento y selección de examinadores, hasta los procedimientos involucrados en la aplicación en aula.

En lo que se refiere al reclutamiento, existe un especial cuidado en que los examinadores cumplan con el perfil deseado, y se privilegia a aquellos postulantes que tengan experiencia previa en la aplicación de pruebas SIMCE. El perfil de los candidatos debe, al menos, considerar:

- Edad: sobre los 20 años.
- Manejo básico (nivel usuario) de herramientas computacionales.
- Profesionales universitarios y de instituciones de educación superior (titulados, egresados o estudiantes que cursen tercer año en adelante) de carreras de las ciencias sociales, educación y salud, tales como: Pedagogía Básica o Media, Psicología, Educación Parvularia, Educación Diferencial, Psicopedagogía, Trabajo Social o Enfermería.
- No poseer antecedentes delictivos, no haber sido mal evaluados en aplicaciones SIMCE anteriores ni tener conflicto de interés con el establecimiento donde se aplica la prueba (Agencia de Calidad de la Educación, 2013b: 103-104).

Posteriormente a la selección y capacitación de examinadores se realizan visitas previas a los planteles. Para ello los examinadores realizan las siguientes actividades:

- Solicitar la colaboración al director del establecimiento para generar condiciones óptimas para la aplicación.
- Verificar con el Libro de Clases la información impresa en el Formulario de Lista de Curso y Formulario de Control en caso de aplicación de la prueba en versión para alumnos con Discapacidad Sensorial.
- Verificar los casos premarcados de estudiantes con NEE permanentes o los casos de los estudiantes con discapacidad sensorial en las listas de curso.
- Visitar y motivar al curso asignado.

- Entregar el (los) cuestionario(s) que será(n) devuelto(s) el día de la aplicación respectiva (Cuestionario Padres y Apoderados, y Cuestionario Docentes), sólo para los casos donde la aplicación se realiza en un día. Además, se debe completar el Formulario de Aplicación, en lo referido a la entrega de los cuestionarios antes mencionados (Agencia de Calidad de la Educación, 2013b).

La aplicación en aula está guiada por el Manual de Aplicación SIMCE (documento de uso interno que no se pudo conseguir para este estudio). Durante la aplicación en aula el examinador debe hacer primero una revisión del material recibido, e informar inmediatamente al supervisor cualquier inconsistencia. Luego, debe preparar la sala de clases con los criterios de ordenación de puestos y asignación de cuadernillos, de acuerdo a lo estipulado en el Manual de Aplicación. Posteriormente el examinador debe ingresar a la sala de clases a los estudiantes, dando paso en primer lugar a aquellos con discapacidad sensorial. Luego de verificar la asistencia, debe aplicar los instrumentos correspondientes a cada día de aplicación.

En términos generales se espera que los estudiantes realicen la prueba en un contexto tranquilo y alejado de ruidos externos. Una vez finalizada la aplicación de las pruebas de acuerdo a los tiempos estipulados, los examinadores retiran éstas, realizan una cuadratura y sellan en el aula la caja con las pruebas en su interior para su posterior entrega y despacho a los centros de operaciones.

#### Análisis de los resultados

Desde una perspectiva de atención a la diversidad, el operativo SIMCE contempla algunas acciones que tienden a contextualizar los resultados de aprendizaje.

En principio y como análisis meramente confirmatorio, se verifica la presencia (o ausencia) de DIF. Cabe recordar que durante el pilotaje también se realiza este tipo de análisis; sin embargo, la ACE suele llevar estos procedimientos después de la aplicación censal para que el proceso sea aún más cuidado. Los DIF se realizan tomando como variables de comparación el género y la ruralidad, utilizando la misma metodología empleada en los análisis de los pilotajes (comparación del parámetro de dificultad). Históricamente en el análisis de las pruebas SIMCE definitivas y experimentales se ha considerado como indicador de DIF una diferencia en dificultad, en valor absoluto, igual o mayor a 0.4.

Ahora bien, en el análisis de los resultados de logro, la ACE reconoce que éstos se encuentran determinados por múltiples factores internos y externos a los establecimientos. La calidad del desempeño del docente, un buen equipamiento o una gestión adecuada del director son variables que, de alguna forma, resultan controlables por las escuelas. En cambio, la condición socioeconómica de los estudiantes o el nivel educacional de sus padres no pueden ser modificados por el establecimiento. Por tanto, para establecer que una comunidad educativa es más efectiva que otra y afirmar, con cierto grado de certeza, que las diferencias de sus resultados en SIMCE se explican por lo que sucede al interior del establecimiento y no fuera de éste, resulta razonable comparar establecimientos que atienden a estudiantes de características socioeconómicas similares.

Para ello se divide a los centros escolares en cinco grupos socioeconómicos a fin de controlar el efecto de este factor al momento de realizar juicios sobre el logro educativo: bajo, medio-

bajo, medio, medio-alto y alto. Las variables utilizadas para realizar esta clasificación son el nivel educacional reportado por los padres y madres de los estudiantes, el ingreso mensual declarado por ellos y el Índice de Vulnerabilidad del Establecimiento. Por la importancia que juega este factor en la interpretación y el uso de los resultados, su construcción se realiza a partir de una metodología que fue diseñada recuperando los avances más recientes sobre el campo y que ha sido transparentada desde su inicio (Agencia de Calidad de la Educación, 2013c).

Así, los grupos socioeconómicos como un elemento de diversidad, están considerados en la mayor parte de las fases de estas evaluaciones, de manera especial en el análisis de la información, la interpretación de los resultados y el uso que se hace de ellos.

En lo referente a género, la ACE ha construido el Indicador Equidad de Género, el cual evalúa el logro equitativo de resultados de aprendizaje obtenidos por hombres y mujeres en establecimientos mixtos. Se calcula sobre la base de los resultados de los establecimientos en las pruebas SIMCE de Comprensión de Lectura y Matemática, y considera la diferencia entre los promedios obtenidos por mujeres y hombres en las últimas tres mediciones, si la prueba es anual; y las últimas dos mediciones, si la prueba es bianual (Leiva, Saavedra y Muñoz, 2013).

#### Uso de los resultados

La ACE genera anualmente informes específicos destinados a distintas audiencias entre los que destacan los padres de familia, los centros escolares, las autoridades, investigadores y académicos del ámbito educativo. En dichos informes se incluye información que permite contextualizar los resultados al menos por el grupo socioeconómico en el que se ubica la escuela, y en otros casos también se proporciona información contextualizada por género, ruralidad y tamaño del plantel escolar.

Los resultados de las pruebas SIMCE tienen diversos usos, tanto de bajo como de alto impacto. Entre los primeros, los resultados permiten entregar retroalimentación a los actores del sistema educativo, aportando información clave para que cada comunidad educativa reflexione sobre los aprendizajes alcanzados por sus estudiantes e identifique desafíos y fortalezas que contribuyan a la elaboración o reformulación de estrategias de enseñanza orientadas a mejorar los aprendizajes. En este sentido, la ACE ha generado el programa Uso de la Información. Este programa pretende fortalecer las capacidades de reflexión colectiva y contextualizada de los aprendizajes, proveer de recursos pedagógicos para el análisis, la reflexión y el desarrollo de acciones de mejora, promover la vinculación de la evaluación interna con la evaluación externa y generar la implicación activa de todos los actores de la comunidad escolar en la evaluación.

Por otro lado, los resultados de las pruebas sirven, junto con los OICE, para crear el Ordenamiento de Establecimientos, que, tomando en cuenta tanto las características de las familias de los estudiantes como su grupo socioeconómico, clasifica a los planteles en cuatro categorías de desempeño (alto, medio, medio-bajo e insuficiente), lo que a su vez tiene consecuencias sobre la cantidad y tipo de apoyos que se proporciona a los establecimientos, la cantidad de visitas de evaluación que reciben y, eventualmente, el cierre del plantel escolar. Asimismo, los resultados contextualizados sirven para que los padres de familia seleccionen el plantel al que desean que sus hijos acudan, a partir de un sistema de *vouchers* que opera en Chile desde los años 80.

Finalmente, es importante señalar que existen algunos estudios (Martín y Alessandri, 2010; Matear, 2007) que, aprovechando las bases de datos de SIMCE, han hecho un análisis del rendimiento educativo en el cual se considera el nivel socioeconómico de los estudiantes, lo que muestra un avance en la consideración de la diversidad para contextualizar los resultados de logro.

### 1.3.2. Pruebas SIMCE Discapacidad Sensorial

Hasta antes de 2008, los estudiantes con alguna NEE permanente habían sido excluidos de los procesos evaluativos de gran escala realizados en Chile. A partir de esa fecha, y derivado de cambios en la legislación, inició el diseño de evaluaciones estandarizadas de logro con un espíritu de inclusión. Como señalan Thompson y Thurlow (citados por la Agencia de Calidad de la Educación, 2013b: 41): “La inclusión de los estudiantes con discapacidad a este tipo de evaluaciones permite conocer más exactamente el desarrollo educacional del país, beneficiar a toda la población estudiantil con las reformas existentes en esta materia, realizar comparaciones más exactas, evitar consecuencias no deseadas producidas por la exclusión, promocionar altas expectativas de logro y fomentar el acceso de todos los estudiantes al currículo general”.

En 2008 el MIDE UC, un centro de investigación, desarrollo y servicios de la Escuela de Psicología de la Pontificia Universidad Católica, por encargo del MINEDUC, hizo un estudio piloto acerca de la accesibilidad de las pruebas SIMCE a estudiantes con discapacidad visual total o parcial y déficit auditivo. Este piloto permitió que a partir de 2009 y hasta 2012, estudiantes con NEE de 4° básico en tres regiones geográficas (Metropolitana, Valparaíso y Biobío) rindieran las pruebas censales en las asignaturas de Lenguaje y Comunicación, para la habilidad de Comprensión de Lectura y de Matemática. A partir de 2013, esta prueba se aplica de forma censal en todas las regiones del país (Agencia de Calidad de la Educación, 2013a; Unidad de Currículum y Evaluación, 2011a).

Para posibilitar la participación de los estudiantes con discapacidad sensorial en las evaluaciones de logros de aprendizaje, fue preciso implementar cambios en los materiales o en los procedimientos de la prueba con el fin de dar accesibilidad a los estudiantes con necesidades especiales y, al mismo tiempo, conseguir la medición del mismo dominio de las pruebas regulares y no las características asociadas a la discapacidad. Estos cambios se denominan *acomodaciones* y su objetivo es permitir la participación autónoma de estudiantes con discapacidad sensorial, conservando los contenidos evaluados del nivel y eliminando las barreras de acceso existentes causadas por su discapacidad.

Las acomodaciones permiten a los estudiantes con discapacidad sensorial enfrentarse a los instrumentos de evaluación en condiciones similares a las de los estudiantes que rinden las pruebas regulares, asegurando su participación en la evaluación. Un cambio en el test es considerado una acomodación si es que no se altera el constructo medido, basado en una necesidad individual, además es efectivo para estudiantes que necesitan este cambio y no lo es para otros estudiantes.



## Diseño de las pruebas

Desde el 2008 con la realización del estudio piloto se definió un equipo de profesionales especialistas en evaluación de estudiantes con NEE (Equipo NEE), para llevar a cabo las acomodaciones correspondientes a las pruebas SIMCE.

Estas pruebas solamente incorporan preguntas de selección múltiple, pero conservan las habilidades y contenidos evaluados así como la dificultad presente en las pruebas regulares SIMCE, según los datos psicométricos que tiene cada pregunta y el juicio experto al que fueron sometidas tanto por profesionales del área de accesibilidad como por los especialistas de las disciplinas evaluadas. Las preguntas de la prueba abarcan distintos grados de dificultad con el fin de recoger información para clasificar y describir el desempeño de los estudiantes.

Para evaluar los ejes y habilidades de Comprensión de Lectura y Matemática, las preguntas y textos presentes en las evaluaciones SIMCE para discapacidad sensorial fueron acomodados según las características presentes en cada discapacidad, considerando los procedimientos descritos a continuación.

El proceso de selección de preguntas inició con la determinación del nivel de accesibilidad que presentaban los reactivos de la prueba SIMCE regular. Ello implicó realizar una preselección de las preguntas, priorizando aquellos ítems que requerían pocas o pequeñas acomodaciones, lo que minimiza las diferencias presentes entre preguntas con y sin acomodación. La preselección se hizo con base en bibliografía especializada, estudios realizados de manera interna y el conocimiento acumulado sobre el funcionamiento de las preguntas y textos durante las evaluaciones anteriormente realizadas a estos estudiantes.

Basados en tales antecedentes, se clasificaron las preguntas y textos considerando la accesibilidad y las acomodaciones necesarias, cuidando siempre que éstas no modificaran el constructo por medir, a través de los siguientes lineamientos (Agencia de Calidad de la Educación, 2013b):

- Presencia de palabras esenciales para responder el ítem.
- Longitud del texto mínimo y organización lo más clara y secuenciada posible.
- Vocabulario y estructura de la oración, apropiados al nivel evaluado.
- Enunciados escritos de manera directa y con planteamientos en estilo positivo.
- Instrucciones claras, mínimas en longitud, redactadas en enunciados en estilo positivo y voz activa.
- Objetivo de evaluación evidente.
- Inclusión de una imagen o elemento visual necesario para responder el ítem.
- Imágenes que representan claramente la intención de la instrucción, cuyo diseño sea presentado lo más simple posible.
- Alternativas breves presentadas de manera clara.
- Alternativas equilibradas respecto a su longitud, orden y contenido, todas ellas plausibles y sólo con una respuesta correcta.
- Diseño organizado de manera coherente y clara.
- Todos los elementos del texto legibles.
- Temática comprensible para la población evaluada.
- Selección de textos de la prueba regular, que obtuvieron una mayor cantidad de preguntas.

- Selección de preguntas de la prueba regular, que por su longitud puedan ser diagramadas en macrotipo y braille.
- Selección de textos que eviten lenguaje figurado y preguntas focalizadas en habilidades metalingüísticas.

Además, hasta el año 2011, las preguntas y textos contaban con revisiones de accesibilidad externas por parte de docentes especialistas que se encontraban en ejercicio con estudiantes con discapacidad sensorial, los cuales fungían como expertos. En posteriores ciclos sólo alrededor de la mitad de los reactivos se han sometido a esa revisión.

En 2011 y 2012 el equipo encargado del diseño de las pruebas desarrolló un par de estudios que buscaban obtener información acerca del funcionamiento de las modificaciones en los reactivos para las tres subpoblaciones a las que estaban dirigidos. El primero tenía como objetivo conocer la forma en que los estudiantes con discapacidad sensorial accedieron a las preguntas y se utilizó la técnica de Pensamiento en Voz Alta en estudiantes con debilidad visual total (DVT) y déficit auditivo (DA). Conjuntamente, se realizaron entrevistas semiestructuradas a docentes para identificar barreras de acceso presente en estas poblaciones. El segundo estudio, de tipo exploratorio, abordó qué formato era más accesible para que los estudiantes pudieran responder el Cuestionario para Estudiantes, ya que esa fue la primera aplicación que se realizó a esta población. Adicionalmente, se aplicó el Test de Velocidad Lectora a estudiantes con DVT, con el objetivo de identificar más exactamente el tiempo promedio utilizado para leer preguntas en braille.

Considerando los criterios y los resultados de las investigaciones detallados anteriormente, se procedió a generar una primera propuesta de acomodaciones a los textos y preguntas seleccionados para armar los instrumentos de evaluación.

Las acomodaciones que finalmente se incorporaron en estas pruebas fueron: destacar palabras o símbolos; omisión o ajuste de imágenes; ajustes en los textos conservando los contenidos evaluados, tales como sustitución léxica; incorporación de información de contexto; incorporación de información instruccional; incorporación de léxico para hacer más comprensible el contexto; omisión o elipsis de información; explicitación de información; descripción de información auditiva; formato de la prueba en braille, imágenes táctiles y formato de la prueba en macrotipo.

#### Aplicación de las pruebas

Previo a la aplicación de las pruebas, se hizo un censo de la cantidad de alumnos que asistiendo a escuela regular o especial, presentaran discapacidad sensorial.

En lo relativo a la aplicación en campo, para procurar que los estudiantes con discapacidad sensorial contaran con igualdad de oportunidades para enfrentar la evaluación, en la capacitación a aplicadores se les informó sobre las acomodaciones permitidas durante la evaluación, las cuales fueron: a) acomodaciones de presentación, tales como uso de dispositivos de aumento, lenguaje de señas, formato de la prueba en braille, imágenes táctiles y formato de la prueba en macrotipo; b) acomodaciones de respuesta, entre las que se incluye escribir respuestas en braille, uso del ábaco y responder en el cuadernillo de prueba, a diferencia de la aplicación regular cuyos alumnos responden en una hoja de respuestas; y c) acomodaciones del contexto de aplicación

como la ubicación del estudiante en la primera fila o donde disponga de las mejores condiciones de iluminación y espacio para aplicar la prueba.

### Análisis de los resultados

Después de la aplicación de las pruebas, se procede a su lectura automatizada, la cual también contempla especificaciones especiales para la prueba de DS. Una vez que se dispone de la base de datos validada inicia el análisis de la información.

Dado el número relativamente reducido de casos, los resultados de la prueba DS se analizan mediante la Teoría Clásica de los Test. Así, las estimaciones se generan en función del porcentaje de aciertos que tuvieron los alumnos con DS (Agencia de Calidad de la Educación, 2013a; Unidad de Currículum y Evaluación, 2012a).

Se tiene la precaución de que, cuando se desea hacer comparaciones con los alumnos que responden las SIMCE regulares, se calculan también los puntajes de estos últimos en función de la proporción de aciertos.

### Difusión de resultados

La ACE genera un informe específico para dar a conocer los resultados de la prueba para alumnos con discapacidad sensorial (Agencia de Calidad de la Educación, 2013a) que está dirigido a docentes y directivos de los centros escolares.

El informe desglosa los resultados por: los dos dominios y ejes evaluados (comprensión lectora, cuyos ejes de habilidad son localizar, interpretar y relacionar, y reflexionar; y matemáticas, cuyos ejes de contenidos son números, geometría y datos y azar, y cuyos ejes de habilidad son conocimiento, resolución de problemas y razonamiento), el tipo de discapacidad (visual total, visual parcial y déficit auditivo) y el tipo de establecimiento al que asisten los alumnos (regulares o especiales).

En el informe de resultados se especifican las habilidades evaluadas así como ejemplos de algunas tareas que se les solicitaron a los alumnos con DS. Asimismo, se informa sobre las acomodaciones utilizadas en el diseño, aplicación y calificación de la prueba.

De la misma manera, se ofrecen orientaciones sobre los casos en que es posible realizar comparaciones de los resultados de estas pruebas. Por ejemplo, en esta prueba se utiliza una escala distinta a la de las pruebas regulares para evitar la tentación de realizar comparaciones. En lugar de usar la escala con media de 250 puntos, aquí se emplea el porcentaje de respuestas correctas, lo cual también está relacionado con que en estos instrumentos se empleó la Teoría Clásica del Test, a diferencia de la TRI que es empleada en las pruebas SIMCE regulares (Agencia de Calidad de la Educación, 2013a).

Se advierte que las puntuaciones conseguidas entre alumnos con distintos tipos de discapacidad no son equivalentes, por lo que no es adecuado hacer comparaciones. A diferencia de lo apropiado que resultaría comparar los resultados entre comprensión lectora y matemáticas al interior de un mismo tipo de prueba.

#### 1.4. OPINIONES EN LOS MEDIOS

En los medios electrónicos de información chilenos (los consultados para este estudio), existe muy poca presencia de las pruebas SIMCE DS. El contenido principal de las escasas notas periodísticas localizadas se refiere al operativo de aplicación de las pruebas; la alusión a los resultados es prácticamente inexistente. En cambio, sobre las pruebas SIMCE regulares existe una considerable producción de notas informativas y de opinión. En los meses recientes al periodo en que se realizó este estudio se identificó una intensa discusión en torno al uso de los resultados de las pruebas para realizar *rankings* de escuelas, aun considerando el nivel socioeconómico. Una buena parte de las opiniones encontradas está a favor de que se sigan haciendo públicos los ordenamientos de los establecimientos, a raíz de la reciente decisión del MINEDUC de restringir los resultados por centro a las comunidades escolares.

### 2.1. DESCRIPCIÓN DEL SISTEMA EDUCATIVO NORTEAMERICANO

En los Estados Unidos no existe un sistema nacional de educación; es decir, las escuelas no son administradas por el gobierno federal, sino que su control recae en cada uno de los distritos escolares, ya que se trata de un sistema altamente descentralizado.

La educación básica comprende primaria y secundaria, conocida como K12, es gratuita desde el 1° al 12° grado, y está organizada de la siguiente forma:

- La educación primaria o *elementary* cubre del 1° al 4° o 7° grado, dependiendo de las políticas estatales o distritales de educación, y los niños la inician a los 6 años de edad.
- La escuela media atiende a los preadolescentes y a los adolescentes entre el 5° y el 9° grado, la mayoría de las veces abarca del 6° al 8° grado.
- El rango superior de la escuela media (del grado 7° al 9°) a veces es llamada *junior high school*. La escuela secundaria o *high school* abarca a los estudiantes de los grados avanzados (generalmente del 9° al 12° grado).

Una vez que el estudiante ha terminado este ciclo, puede avanzar a lo que se conoce como estudios terciarios. En estos se incluyen los estudios vocacionales o técnicos (con duración de 2 años) y estudios profesionales o universitarios (4 años). Cuando se ha completado un grado universitario, los alumnos pueden avanzar a estudios de posgrado (maestría, doctorado, programas de posdoctorado).

Estados Unidos tiene una larga tradición en materia de evaluación educativa. Sobre este país se recuperaron las acomodaciones o adecuaciones derivadas de la ley *No Child Left Behind*, así como prácticas evaluativas de atención a la diversidad de dos organismos de evaluación y monitoreo.

### 2.2. NO CHILD LEFT BEHIND ACT

La Ley de Educación Primaria y Secundaria (ESEA, por sus siglas en inglés) es la ley federal más importante de los Estados Unidos relacionada con la educación, desde preescolar hasta el nivel de secundaria.

En su más reciente reautorización por parte del Congreso estadounidense, la ESEA se conformó en 2001 como la ley *No Child Left Behind* (NCLB); el propósito descrito en la Sección 1001 del Capítulo I, "Improving the Academic Achievement of the Disadvantaged" es asegurar, mediante varias acciones, que todos los niños tengan oportunidades justas e igualitarias de recibir educación de alta calidad y alcanzar el dominio mínimo en los estándares de logro académicos y en las evaluaciones académicas estatales (Cortiella, 2005). Dichas acciones incluyen:

1. Asegurar que las evaluaciones académicas, sistemas de rendición de cuentas, preparación y formación de docentes, planes de estudio y materiales educativos cumplan con las normas académicas del estado para que los estudiantes, profesores, padres y administradores puedan medir el avance en función de expectativas conocidas.
2. Satisfacer las necesidades educativas de los niños de bajo rendimiento en las escuelas de mayor pobreza, que muestran dominio limitado de inglés, niños migrantes, niños con discapacidades, niños indígenas, niños en riesgo de delincuencia, niños con requerimiento de asistencia lectora.
3. Cerrar la brecha en el rendimiento entre los grupos de estudiantes minoritarios y no minoritarios, y entre los estudiantes desfavorecidos y los más favorecidos.
4. Revertir el bajo rendimiento de las escuelas que no han logrado proporcionar una educación de alta calidad a sus estudiantes, al tiempo de proporcionar alternativas a los estudiantes en escuelas de este tipo que les permitan recibir una educación de alta calidad.
5. La mejora y el fortalecimiento de la rendición de cuentas, la enseñanza, y el aprendizaje mediante el uso de sistemas estatales de evaluación diseñados para asegurar que los estudiantes alcancen los estándares de logro estatales y aumentar el rendimiento en general, pero especialmente el de los grupos más desfavorecidos.
6. La distribución y orientación de los recursos en las escuelas cuyas necesidades sean mayores para lograr hacer una diferencia.
7. Proporcionar una mayor autoridad y flexibilidad en la toma de decisiones a las escuelas y los maestros a cambio de una mayor responsabilidad sobre el desempeño de los estudiantes.
8. Elevar la calidad de la enseñanza a través de proporcionar abundantes oportunidades de desarrollo profesional a los profesores, la coordinación entre los servicios educativos y, en la medida de lo posible, con otros organismos que prestan servicios a la población escolar.
9. Proporcionar oportunidades a los padres de familia para participar en la educación de sus hijos.

En sus diversos apartados la Ley asigna fondos y propone estrategias para mejorar los resultados de logro, así como para atender a la diversidad tanto en el aula como durante las evaluaciones nacionales. Destaca la asignación de recursos para contratación y capacitación de profesores y directores, ya que se pretende incrementar el logro académico de los alumnos con la implementación de diversas estrategias; por ejemplo, a través de la mejora de la calidad de los profesores y de los directores de escuelas, y el incremento del número de profesores, directores y asistentes de la dirección altamente calificados, para que las escuelas muestren resultados y sean responsables de la mejora académica de los alumnos.

Se plantea también la asignación de fondos para atender a las escuelas a las que asisten niños de familias de bajos recursos económicos. La elegibilidad de estos niños es determinada por la agencia local educativa para otorgar fondos a las escuelas bajo la sección 1113(C)(1), la cual busca proporcionar servicios educativos complementarios que sean consistentes con el programa escolar y los estándares académicos, en la atención a estudiantes con discapacidad, a estudiantes de las principales razas y grupos étnicos, y a niños con competencia limitada en inglés.

Sobre esta última condición, a fin de que los alumnos puedan alcanzar los mismos estándares de logro académico, se busca promover sistemáticamente programas educativos orientados a los niños con competencia limitada en inglés. Esto incluye promover el desarrollo de habilidades de la lengua y comprensión multicultural, proporcionar asistencia a los indígenas americanos con ciertas modificaciones relativas a su estatus de indígenas bajo la ley federal, desarrollar y

repartir materiales de investigación y asistencia técnica enfocada en la escuela, y desarrollar programas que fortalezcan y mejoren la capacitación profesional del personal que trabaja con los niños en estas condiciones.

En este contexto todos los estudiantes de escuelas públicas en todo el país deben participar en pruebas anuales en las áreas académicas específicas y de calidad señaladas en la Ley, incluyendo a los estudiantes con discapacidades. La inclusión de todos los estudiantes en las evaluaciones ayuda a garantizar que las escuelas, distritos escolares y estados sean responsables del logro de estos estudiantes; los resultados de las pruebas son la fuente de información primaria de logro de los estudiantes, y las escuelas son las entidades responsables del desempeño de los estudiantes (Davey, 1992).

### 2.2.1. Atención a la diversidad en las evaluaciones

La evaluación en los Estados Unidos de Norteamérica muestra una práctica sistemática de atención a la diversidad que se refleja en la preocupación por incluir a todos los grupos de estudiantes que asisten a las escuelas (August y McArthur, 1996; GACE, 2014; Office of Learning/ Instruction, Standards, Assessment, and Accountability Unit, 2014; Oklahoma State Department of Education, s/f). A continuación se describe cómo se concibe la atención a la diversidad en dos evaluaciones educativas nacionales; en un caso la inclusión se refiere a alumnos con discapacidad, y en otro a los estudiantes con una competencia limitada de inglés. En ambos casos subyace la preocupación por maximizar su accesibilidad a las evaluaciones.

### 2.2.2. Evaluación educativa de los alumnos con discapacidad

De acuerdo a la ley NCLB, los estudiantes con discapacidades tienen derecho a contar con servicios especiales (acomodaciones) adecuados para participar en las pruebas nacionales de logro. Determinar las adaptaciones que los estudiantes con discapacidades necesitan para poder participar de manera plena e igualitaria en las pruebas de gran escala es un componente fundamental del programa individualizado de educación (IEP, por sus siglas en inglés) del estudiante. De acuerdo a la Guía para el Programa Individualizado de Educación (Kupper, 2000), los miembros del equipo del IEP, incluyendo a los padres, deben involucrarse en un proceso reflexivo que permita determinar las condiciones especiales necesarias para facilitar el acceso del alumno a la instrucción del nivel educativo y su plena participación en las evaluaciones del estado y del distrito escolar. La definición del programa individualizado del estudiante sigue un proceso en el que primero se identifica al alumno que requiere educación especial y servicios asociados, y luego es evaluado en áreas relacionadas con la discapacidad para decidir si es elegible para recibir educación especial. En esta etapa, si los padres del alumno discrepan de los resultados de la evaluación, ellos pueden solicitar otra evaluación independiente a la escuela. Una vez que se decide la elegibilidad del alumno con algún tipo de discapacidad, el equipo del IEP se reúne para escribir el plan individualizado de educación del niño y se realiza su implementación. El siguiente paso es medir el progreso del alumno, información que es reportada a los padres de familia en los mismos términos de tiempo en que se reporta el progreso de los alumnos que no presentan discapacidad. El plan individualizado de educación se revisa una vez al año o más, a solicitud de los padres de familia (en esta revisión participan en el equipo los padres de familia). Finalmente, el alumno es reevaluado al menos cada tres años con el propósito de definir si continúa siendo un alumno con discapacidad e identificar sus necesidades. Esta reevaluación puede ser realizada antes de los tres años de ser necesario.

Es claro que el IEP es un documento importante para los alumnos con discapacidad, de su implementación se esperan mejoras en la enseñanza, el aprendizaje y los resultados.

### 2.2.3. ¿Qué son los servicios especiales (acomodaciones)?

Los servicios especiales o acomodaciones para estudiantes con discapacidad son herramientas y procedimientos que proporcionan igualdad de acceso a la instrucción y a la evaluación a los estudiantes con discapacidades. Se proporcionan para “equilibrar el campo de juego”. Sin los servicios especiales o acomodaciones, los estudiantes con discapacidades no podrían acceder a la educación ni participar plenamente en las evaluaciones, y tampoco podrían demostrar sus talentos sin que los puntajes que se obtuvieran dependieran de factores distintos a los que las pruebas pretenden medir.

### 2.2.4. Acomodaciones vs. modificaciones

Las acomodaciones o servicios especiales no son lo mismo que las modificaciones, por lo tanto es importante hacer una diferenciación. Las acomodaciones pretenden disminuir los efectos de la discapacidad del estudiante; no pretenden reducir las expectativas de aprendizaje. Una modificación o alteración generalmente implica cambiar, disminuir o reducir las expectativas de aprendizaje respecto a los estudiantes con discapacidad.

A diferencia de los servicios especiales o acomodaciones, el uso constante de las modificaciones puede aumentar la brecha entre el logro de los estudiantes con discapacidades y las expectativas sobre el nivel de logro. Esto puede tener un impacto negativo en la trayectoria escolar del estudiante a tal grado que se pondría en riesgo su avance y la conclusión de los estudios.

### 2.2.5. Acomodaciones para evaluación frente a acomodaciones para enseñanza

De manera ideal, las acomodaciones o servicios especiales en las instalaciones donde se realizan las evaluaciones del distrito o del estado, deberían ser iguales o similares a los que se dan en el salón de clases. Sin embargo, es importante señalar que algunas de las acomodaciones o servicios especiales son exclusivos para la enseñanza y no pueden ser utilizados durante las evaluaciones del distrito o del estado.

### 2.2.6. Acomodaciones o servicios especiales estándar frente a no estándar

Las acomodaciones o servicios especiales estandarizados para personas con discapacidad son aquellos que permiten que se administre la evaluación y se dé la enseñanza, y que no cambian la habilidad que está siendo evaluada. Una acomodación no estandarizada cambia la naturaleza de la tarea o habilidad que se pretende evaluar. Por ejemplo, leer una prueba de lectura en voz alta a un estudiante cuando se pretende evaluar esta habilidad se considera una acomodación no estándar porque no resultará en una medición real de la habilidad de decodificación; si, por otro lado, el test está midiendo lectura de comprensión, la lectura del texto no cambiará la habilidad objetivo y permitirá demostrar la habilidad de comprensión sin la interferencia de la discapacidad.



Las políticas relacionadas con acomodaciones o servicios especiales estandarizados y no estandarizados varían de estado a estado, todos los miembros del IEP necesitan conocer las políticas del estado en relación con el uso de las acomodaciones.

Para poder tomar decisiones razonables acerca del tipo de acomodaciones pertinentes de acuerdo con los usuarios, se requiere conocer:

- La prueba: su contenido, tipos de preguntas y condiciones de la evaluación
- Los lineamientos estatales de evaluación
- Los lineamientos estatales de acomodación
- Qué acomodaciones pueden invalidar una evaluación

### 2.2.7. Selección de las acomodaciones

La premisa es que todos los estudiantes tienen derecho a acomodaciones apropiadas que les permitan participar de manera completa en las evaluaciones a gran escala de sus estados y distritos escolares.

De acuerdo a la iniciativa NCLB la selección de las acomodaciones tanto para la enseñanza como para la evaluación recae en el IEP estatal y se hace en función de lo que cada estudiante necesita, y no con base en la categoría de la discapacidad. Una vez que se ha seleccionado el tipo de acomodación, ésta deberá ser utilizada de manera sistemática para la instrucción y la evaluación. Cada profesor y los responsables de las instalaciones deben ser informados de las acomodaciones o servicios especiales que deben proporcionarse.

Las siguientes son algunas preguntas que se debe hacer el equipo del IEP para cada estudiante al momento de seleccionar los tipos de acomodaciones:

- ¿Cuáles son las fortalezas y necesidades de aprendizaje del estudiante?
- ¿Cómo es que las necesidades de aprendizaje del estudiante influyen en los estándares de logro de los contenidos del grado escolar?
- ¿Qué instrucción especializada necesita el estudiante para alcanzar los estándares de logro de los contenidos del grado escolar?

Una vez que se ha dado respuestas a estas preguntas, se sugiere revisar y discutir los tipos de acomodaciones proporcionadas al estudiante a través de las siguientes preguntas:

- ¿Qué acomodaciones ha estado utilizando el estudiante de manera regular tanto en el salón de clase como a la hora de tomar exámenes?
- ¿Cuál es la percepción de los estudiantes sobre la funcionalidad de esas acomodaciones?
- ¿El estudiante ha estado dispuesto a utilizar esas acomodaciones o servicios especiales?
- ¿Cuál es la percepción de padres de familia, maestros y otras personas sobre cómo han funcionado esas acomodaciones o servicios especiales?
- ¿Ha habido dificultades para proporcionar las acomodaciones que se han seleccionado?

Cuando se haya decidido implementar un nuevo tipo de acomodación es importante prever cuánto tiempo le llevará al estudiante aprender bajo ese nuevo servicio especial y cuándo es que se llevará a cabo ese aprendizaje. Es importante también asegurar que haya suficiente tiempo para

que el estudiante aprenda a utilizar el tipo de acomodación o de servicio especial antes de que sea parte regular de la aplicación de la evaluación a gran escala del estado y del distrito escolar.

La evaluación de la efectividad de las acomodaciones debe ser un proceso permanente. Sólo con una revisión cercana de su impacto se pueden implementar mejoras. Las revisiones anuales deberán incluir:

- Las acomodaciones utilizadas y los resultados en las evaluaciones cuando fueron utilizadas las distintas acomodaciones.
- La percepción de los estudiantes sobre qué tan bien ha funcionado la acomodación.
- Combinaciones de acomodaciones que han probado ser efectivas.
- Las percepciones de los maestros, paraprofesionales y otros profesionales sobre la funcionalidad de las acomodaciones.

La ley NCLB diseñó una guía para seleccionar la acomodación más adecuada de acuerdo a la discapacidad. La tabla 9 muestra esta guía en la que se identifican los posibles beneficiarios, las preguntas que se debe hacer para la toma de decisiones y ejemplos de acomodaciones.

**Tabla 9** Guía para elegir acomodaciones o servicios especiales

Beneficiarios	Preguntas	Ejemplos
<b>Acomodaciones o servicios especiales en presentaciones</b>		
Estudiantes con discapacidad para la lectura, definida como dificultad o incapacidad para leer formas impresas estándar debido a una discapacidad física, sensorial o cognitiva.	<ul style="list-style-type: none"> <li>• ¿El estudiante puede leer y entender instrucciones?</li> <li>• ¿El alumno necesita que se le repitan instrucciones de manera frecuente?</li> <li>• ¿El estudiante ya ha sido reconocido con una discapacidad para la lectura?</li> </ul>	<ul style="list-style-type: none"> <li>• Tipografía grande</li> <li>• Aparatos o instrumentos que amplifiquen los objetos</li> <li>• Una persona que lea para ellos</li> <li>• Grabaciones de audio</li> <li>• Lector de pantalla</li> <li>• Materiales parlantes (calculadoras, relojes, etc.)</li> </ul>
<b>Acomodaciones o servicios especiales para brindar respuestas</b>		
Estudiantes con discapacidades físicas, sensoriales o de aprendizaje (incluyendo dificultades con la memoria, secuencias, dirección, alineación y organización).	<ul style="list-style-type: none"> <li>• ¿Puede el estudiante utilizar un lápiz u otro instrumento de escritura?</li> <li>• ¿El estudiante tiene alguna discapacidad que le impide deletrear?</li> <li>• ¿El estudiante tiene problemas con el seguimiento de una página a otra y mantener su lugar?</li> </ul>	<ul style="list-style-type: none"> <li>• Transcripción-copia de textos</li> <li>• Toma de notas</li> <li>• Grabaciones</li> <li>• Respuestas en cuadernillos de examen</li> <li>• Aparatos o dispositivos para ortografía y gramática</li> <li>• Organizadores gráficos</li> </ul>
<b>Acomodaciones o servicios especiales - Programaciones, extensión de tiempo</b>		
Estudiantes que necesitan más tiempo, no se pueden concentrar por periodos largos de tiempo, tienen discapacidades relacionadas con la salud, se cansan fácilmente, tienen una dieta especial y tienen necesidades de toma de medicamentos.	<ul style="list-style-type: none"> <li>• ¿Pueden los estudiantes trabajar de manera continua durante todo el tiempo que dura el examen?</li> <li>• ¿El estudiante se cansa fácilmente debido a problemas de salud?</li> <li>• ¿El estudiante necesita descansos cortos y frecuentes?</li> </ul>	<ul style="list-style-type: none"> <li>• Extensión del tiempo</li> <li>• Descansos frecuentes</li> <li>• Múltiples sesiones de evaluación</li> </ul>
<b>Acomodaciones de los espacios</b>		
Estudiantes que se distraen fácilmente en grupos grandes y que se concentran mejor en grupos pequeños.	<ul style="list-style-type: none"> <li>• ¿Los estudiantes se distraen con facilidad con los otros alumnos?</li> <li>• ¿Los estudiantes tienen problemas para completar una tarea?</li> <li>• ¿Los estudiantes muestran una conducta que puede causar interrupciones en los otros alumnos?</li> </ul>	<ul style="list-style-type: none"> <li>• Cambiar al estudiante de salón</li> <li>• Utilización de audífonos o tapones para los oídos</li> <li>• Cubículos</li> </ul>

### 2.2.8. Estudios de validez

El panel de expertos para estudios de validez de la Evaluación Nacional del Progreso Educativos (NAEP, por sus siglas en inglés), planeó realizar un trabajo sobre las acomodaciones que se proporcionaban a estudiantes con discapacidad. El estudio pretendía conocer el impacto del aumento del tiempo para contestar las pruebas sobre el desempeño de los estudiantes discapacitados y no discapacitados. Se reconocía la necesidad de conocer el impacto de las acomodaciones en el desempeño de los estudiantes con el fin de saber si los resultados de los estudiantes discapacitados se podían combinar de una manera válida con los resultados de otros estudiantes. Este estudio estaba previsto a ser realizado en 1997 (Olson y Goldstein, 1997), pero al parecer no se concretó.

En 2002, Weston realizó un estudio de validez de servicios especiales o acomodaciones en pruebas de expresión oral. De acuerdo con Weston (2002), la inclusión completa de los estudiantes con discapacidad en las evaluaciones a gran escala ha sido vista como necesaria y positiva, no solamente por asuntos de equidad y accesibilidad, sino también como un medio para mejorar la comparabilidad de las evaluaciones entre las unidades y a través del tiempo. Como ya se mencionó, las acomodaciones o servicios especiales para estudiantes con discapacidad son prácticas que promueven la participación de estos estudiantes en los sistemas de evaluación y de rendición de cuentas.

Una acomodación razonable en las evaluaciones permite el acceso a las pruebas de los estudiantes con discapacidad a través de la modificación en los procesos de administración de la prueba que no cambian la naturaleza del constructo que se quiere medir. Algunos de los servicios especiales o acomodaciones incluyen: extensión del tiempo para contestar la prueba, presentación oral y respuestas orales, intérpretes de lenguaje por señas, braille, cambios en el tamaño del formato, reubicación en salón sin ruido, y ayudas computarizadas. Weston (2002) reconoce que la provisión de acomodaciones es un tema controvertido porque existe poca investigación empírica sobre los efectos de las adaptaciones o acomodaciones específicas, aunque es presumible que, de manera inadvertida, las acomodaciones que alteran las condiciones de la administración de los exámenes puedan cambiar el significado de un resultado, causando interferencias, y que por ende los resultados y las decisiones basados en éstos puedan ser erróneos. De acuerdo con Weston (2002), la investigación empírica puede proporcionar evidencia para evaluar la validez de las acomodaciones que se utilicen.

## 2.3. NATIONAL CENTER FOR EDUCATION STATISTICS (NCES)

En 1996 se dictaron las actas de la conferencia sobre inclusión y acomodo para estudiantes con una competencia limitada de inglés en la NAEP en Washington DC (August y McArthur, 1996).

En esta conferencia se reconoció y enfatizó la necesidad de desarrollar una serie de lineamientos para determinar cómo incluir estudiantes con una competencia limitada de inglés (LEP, por sus siglas en inglés) en la NAEP.

Lo que se buscaba era maximizar el número de estudiantes en la evaluación nacional que pudieran ser evaluados de manera válida, minimizando el número de procesos alternos de evaluación y manteniendo las decisiones fluidas, simples, consistentes y realistas dentro del contexto de

la NAEP. Se tendrían que desarrollar criterios para definir el mejor ajuste entre las características particulares de los estudiantes con LEP y una forma particular de evaluación: una versión sin modificación en inglés, una versión del examen en la lengua nativa del estudiante, una versión sin modificación en inglés pero con apoyo, una versión modificada en inglés, o un número de versiones alternativas. Estas propuestas son coherentes con la idea de justicia en las evaluaciones, desde el diseño de las pruebas, al atenderse la situación real de las diferencias de los alumnos en el desempeño de la lengua.

Una propuesta señalaba que a aquellos estudiantes con LEP que fueran lo suficientemente competentes en inglés se les debería administrar la versión en inglés sin modificaciones. Los criterios de participación se basarían en el nivel de competencia en inglés, en lugar de años de instrucción de inglés u otras características similares. Esto se propuso porque los años de instrucción de inglés no reflejan de manera fiel el desempeño en el idioma, dado que existe una amplia variación de competencias en inglés en casa y en los antecedentes escolares entre los alumnos con LEP (August y McArthur, 1996).

Las modificaciones que se realizaron a la NAEP para hacer más inclusiva la participación de los estudiantes con LEP incluyeron las siguientes acciones:

- Desarrollo de los instrumentos de evaluación en la lengua materna.
- Utilización de inglés estándar en los instrumentos de evaluación con varios tipos de apoyo.
- Modificaciones de la versión en inglés estándar.

De acuerdo al NCES, aproximadamente 73% de los estudiantes con LEP provienen de un antecedente hispanoparlante, de tal suerte que parecía pertinente desarrollar una versión en español de NAEP para ser utilizada con este subgrupo ya alfabetizado en su lengua nativa; sin embargo, se sugirió tener mucha precaución en la traducción al español de la versión original, la identificación de un lenguaje apropiado dadas las distintas variedades de español de los diferentes países, la identificación y minimización de las diferencias culturales y la búsqueda de palabras y frases en español equivalentes a aquéllas en inglés. Por otro lado, la evaluación de estudiantes cuya lengua nativa no fuera español no parecía una opción viable, debido a lo reducido de esos grupos, además de que parecía muy difícil para el NCES conseguir una muestra suficiente que permitiera reportar resultados de la evaluación para cada uno de esos grupos (August y McArthur, 1996).

Las evaluaciones en inglés son difíciles para los estudiantes con LEP porque valoran contenidos y conceptos y la habilidad verbal, de manera particular en las áreas de lectura de comprensión y escritura (Solano-Flores, 2012b). Un decremento en la carga del idioma inglés puede hacer la evaluación de la lengua más apropiada para los estudiantes con LEP. Algunas estrategias alternativas pueden ser divididas en aquéllas que incluyen modificaciones a los ítems y reactivos e instrucciones, con la idea de simplificar la carga de la lengua, y aquéllas que proveen apoyo durante la administración de los ítems sin modificarlos, por ejemplo, clarificación adicional ya sea al final o durante la administración de la evaluación, instrucciones grabadas, disponibilidad de más tiempo de respuesta, etcétera. En todos los casos es importante tener en cuenta las capacidades académicas de los estudiantes cuando se piense en implementar acomodaciones durante las evaluaciones.

Una proporción mensurable de estudiantes con LEP puede ser excluida de la evaluación aun con las modificaciones y el acceso al español. Es importante recuperar la información de esos

estudiantes aunque los datos no alcancen los criterios de validez y confiabilidad de NAEP. Por ejemplo, los puntajes o resultados de NAEP pueden ser asignados a esos estudiantes basados en las calificaciones de los maestros, o en imputaciones basadas en la lengua nativa del estudiante, o antecedentes educacionales de los estudiantes o una combinación de las medidas anteriores. Otras fuentes alternativas de información pueden incluir el uso de portafolios, ampliación de conceptos, provisión de andamiaje y clarificación de instrucciones así como demostraciones. Finalmente se recomendó que se consideraran las necesidades de los estudiantes con LEP durante la construcción y desarrollo de las evaluaciones estandarizadas, tales como un decremento en la demanda del idioma inglés, reducción del número de ítems o reactivos y utilización de instrucciones simplificadas. Estas acomodaciones pueden ser logradas sin comprometer la validez de la evaluación de los estudiantes de habla inglesa.

#### Calificación/valoración

Los participantes en la conferencia subrayaron la importancia de desarrollar rúbricas y procedimientos que fueran apropiados para los estudiantes con LEP, por ejemplo, que consideren sus antecedentes culturales y lingüísticos. También recomendaron estudiar la factibilidad del desarrollo de esos instrumentos.

#### Pruebas de campo

Los participantes concluyeron que se deben realizar pruebas de campo antes de que se implemente un instrumento y recomendaron establecer criterios para determinar cuáles métodos podrían ser sometidos a estas pruebas y cuáles otros requerirían mayor investigación y desarrollo.

### 2.3.1. Reporte de los resultados

Los participantes en la conferencia subrayaron la importancia de una definición estandarizada de LEP. En la actualidad existe mucha variación entre los estados y los distritos escolares sobre la forma en que son identificados los alumnos con estas particularidades, de tal suerte que las mediciones características de LEP no reflejan la misma población entre las diferentes jurisdicciones. De igual manera, la mayoría de los participantes en la conferencia recomendaron que, en el caso de los resultados de aquellos estudiantes con LEP que toman el examen sin acomodaciones o ajustes, el NCES reporte los datos tanto de manera separada como formando parte de otros agregados. Así, idealmente los datos se deberían presentar de tres maneras:

1. Sobre todos los alumnos, incluyendo a los estudiantes con LEP
2. Solamente sobre los estudiantes con LEP
3. Sobre todos los estudiantes, excluyendo a los estudiantes con LEP

Además, se recomendó que se reportaran resultados de los estudiantes con LEP por tipo de acomodación o de ajuste. Existen varios aspectos que están relacionados con la evaluación de este tipo de estudiantes que incluyen la disponibilidad de datos, la participación y la validez de los resultados de la evaluación (Olson y Goldstein, 1997).

La agenda federal sobre inclusión y acomodaciones en las evaluaciones enfatizó la importancia de realizar mayor investigación al respecto. Las preguntas principales a las que se debe dar respuesta incluyeron:

- ¿Cuál es la mejor manera para conceptualizar el desempeño en la lengua inglesa?
- ¿Cuáles son los criterios de desempeño en las diferentes dimensiones, de estudiantes con LEP para que puedan o no tomar una evaluación estandarizada desarrollada en inglés?
- ¿Cómo es que se relacionan los contenidos de los instrumentos de evaluación con los niveles de desempeño en inglés?
- ¿Qué modificaciones se pueden hacer en las evaluaciones a gran escala —tanto en la evaluación misma, como en los procesos para administrar la evaluación— para poder incorporar más estudiantes con LEP?
- ¿Cómo es que esas modificaciones pueden afectar la confiabilidad y la validez de las evaluaciones?
- ¿Cómo se define qué estudiantes con LEP toman qué evaluaciones (antecedentes del estudiante, desempeño en la lengua inglesa, antecedentes escolares)?
- ¿Es posible asignar o imputar calificaciones a estudiantes con LEP con base en información sobre sus antecedentes, tales como nivel de desempeño del lenguaje, historia escolar, logros académicos? Si esto es posible, ¿qué variables de los antecedentes podrán predecir los resultados tanto en NAEP como en un escenario académico?
- ¿Cuál será la mejor manera de reportar los resultados para los estudiantes con LEP dados los problemas metodológicos discutidos?

### 2.3.2. Monitoreo

Es crucial el monitoreo de la exclusión de los estudiantes con LEP, a fin de asegurar que todos quienes estén en condiciones de participar en la evaluación lo hagan. El personal de las escuelas —que es generalmente el que decide cuándo y cómo evaluar a sus estudiantes— debe tener lineamientos claros para la toma de decisiones que permitan la inclusión de los estudiantes con LEP. Se recomienda que una persona por distrito escolar sea quien autorice la exclusión de estudiantes y quien proporcione información adicional de éstos. Realizar estudios de seguimiento sobre estudiantes excluidos puede proporcionar información acerca de procedimientos de evaluación y modificaciones a ser desarrolladas o implementadas.

Finalmente, los participantes recomendaron el establecimiento de un comité para dar asesoramiento permanente al NCES sobre asuntos de evaluación a estudiantes con LEP y para que realice investigación permanente y haga recomendaciones sobre necesidades de investigación en torno al tema.

## 2.4. EDUCATIONAL TESTING SERVICE (ETS)

El Servicio de Evaluación Educativa (ETS, por sus siglas en inglés) desarrolla, administra y califica más de 50 millones de exámenes de manera anual, en más de 180 países y en más de 9 000 localidades en todo el mundo.

El ETS fue fundado en 1947 bajo la premisa fundamental de contribuir al progreso de la educación en los Estados Unidos. La misión del ETS es contribuir al avance en la calidad y la equidad de la educación a través de una norma de evaluaciones justas y válidas, de investigación y de productos relacionados. Un componente central de esta misión es el desarrollo de estándares de calidad y de equidad que fueron adoptados como política corporativa en 1981, y desde

entonces continúan como referentes de esta organización. Las normas fueron revisadas en 1987 y 2000, tras la revisión de los estándares educativos y pruebas psicológicas publicada por la American Psychological Association y el Consejo Nacional de Medición en la Educación. Los dos conjuntos de normas son comparables, difiriendo en áreas que reflejan la naturaleza específica de lo que ofrece el ETS (2014a).

La edición de 1987 de las normas del ETS tenía tres niveles de especificidad: principios, políticas y directrices de procedimiento. Las nuevas normas de ETS tienen dos niveles de especificidad: 1) el objetivo esencial de cada estándar, y 2) el estándar mismo. Los estándares se aplican a todos los productos y servicios desarrollados por el ETS, incluyendo las evaluaciones de lápiz y papel, evaluaciones a través del uso de computadora, portafolios, evaluaciones multimedia, evaluaciones de desempeño u otras formas de evaluación.

Del amplio rango de estándares desarrollados, a continuación se señalan aquellos que consideran la diversidad en las evaluaciones educativas.

#### Idoneidad

Estos estándares buscan asegurar que los servicios y productos del ETS satisfagan el propósito previsto para la población de interés. Estos estándares hacen referencia al concepto de validez y recogen evidencia lógica o empírica de que el producto o servicio es el idóneo para la población en cuestión.

#### Equidad

Los estándares de equidad buscan garantizar que los productos y servicios se diseñarán, desarrollarán y administrarán de manera que se atienda de igual manera a cada individuo independientemente de diferencias en características personales como raza, etnicidad, género o discapacidad, y que no son relevantes para el uso previsto del producto o servicio (tests, exámenes, pruebas).

Existen muchas definiciones de equidad en la literatura, e incluso en ocasiones contradictorias entre sí; en este contexto, equidad se refiere a tratar a la gente con imparcialidad al margen de las características personales señaladas arriba que no interfieren en su interacción con el ETS. Con respecto a las evaluaciones, el concepto de equidad requiere entender que son irrelevantes las características personales de los sujetos que son examinados y que no tienen ningún efecto apreciable sobre los resultados de las pruebas o su interpretación. Para las personas con discapacidad, la igualdad de trato puede implicar proporcionar espacios adecuados a necesidades particulares, para dar igualdad de oportunidades en la utilización y aprovechamiento del producto o servicio —tests, pruebas, exámenes— a todas las personas.

El ETS recibe más de 12 000 solicitudes anuales para acomodaciones para personas con algún tipo de discapacidad; se trata de estudiantes que tomarán alguna prueba de egreso de la *High School* o pruebas profesionales. Las agencias como el ETS son responsables de proporcionar a los candidatos con discapacidad acceso a cualquier test que ellos requieran, a través de un certificado de elegibilidad que tiene dos propósitos principales:

- Verificar el uso de una acomodación o servicio especial de un individuo, ya sea en la escuela o en el lugar de trabajo.
- Como un atajo para la aprobación de ciertas acomodaciones o servicios especiales para una discapacidad en particular.

Las acomodaciones o servicios especiales relacionados con algún tipo de discapacidad que proporciona el ETS (2014b) se presentan en la tabla 10.

**Tabla 10** Acomodaciones o servicios especiales generales

Acomodación	Descripción
Extensión del tiempo de duración del examen (todos los exámenes tienen tiempos asignados)	<ul style="list-style-type: none"> <li>• 50% (el tiempo de duración del examen y medio más)</li> <li>• 100% (doble del tiempo de duración del examen)</li> </ul>
Descansos extra	No se incluyen recesos durante el tiempo del examen, en estos casos se puede utilizar estos recesos para toma de medicamentos, bocadillos, ida al baño, etcétera)
Acomodaciones para exámenes que se administran a través de computadoras	<ul style="list-style-type: none"> <li>• Teclado ergonómico</li> <li>• Teclado con área sensitiva (<i>touchpad</i>)</li> <li>• Ampliación de la pantalla</li> <li>• Colores seleccionables en el fondo y en el primer plano</li> <li>• <i>Trackball</i> para mover el cursor</li> </ul>
Auxilio	<ul style="list-style-type: none"> <li>• Lector humano</li> <li>• Escritura</li> </ul>
Auxilio para instrucciones verbales	<ul style="list-style-type: none"> <li>• Intérprete oral o de lenguaje de signos</li> <li>• Copia impresa de instrucciones orales</li> </ul>
Auxilio para toma de notas	Pizarra y lápiz braille

El ETS también ofrece acomodaciones o servicios especiales denominados “acomodaciones menores relacionados con la salud”. Éstos se refieren a una variedad de condiciones médicas que afectan las actividades cotidianas tales como trastornos digestivos, función inmune, respiración, circulación, funciones endócrinas, etcétera. Las condiciones de salud documentadas incluyen entre otras, diabetes, epilepsia y dolor crónico. Algunas de estas afecciones requieren de servicios especiales o acomodaciones menores que implican, por ejemplo:

- Iluminación especial
- Mesas o sillas ajustables
- Descansos para tomar medicamentos o bocadillos
- Cojines amortiguadores

Existen otros servicios especiales o acomodaciones que no requieren de aprobación:

- Ayudas necesarias para deambular: bastón, muletas, sillas de ruedas, andador, prótesis o animales de servicio.
- Ayudas necesarias para escuchar: audífono o amplificador de la voz.
- Ayudas por razones de salud, como el pulsímetro.

Si la persona utiliza una bomba de insulina, no necesita solicitar algún servicio especial o acomodación, a menos que la bomba sea especialmente ruidosa, en cuyo caso es probable que perturbe a otros examinantes, así que puede programarse su examen en una sala separada.



## 2.5. OPINIONES EN LOS MEDIOS<sup>1</sup>

El tema de la evaluación a gran escala es noticia de interés en la actualidad porque se están haciendo las propuestas presupuestarias del gasto público en educación. Las opiniones sobre las pruebas obligatorias a las que se tuvo acceso en los medios de comunicación electrónica consultados son tanto favorables como desfavorables. Los detractores buscan eliminar las pruebas y reducir los fondos para ciertos programas educativos; por ejemplo, algunos sindicatos de maestros consideran que en la práctica se hace demasiado énfasis en las pruebas sin que eso signifique mejora en el aprendizaje; por otro lado, se reconoce la utilidad de estas herramientas para medir el progreso de los estudiantes y para ayudar a los maestros a evaluar sus métodos y a realizar correcciones en caso de ser necesario.

La presidencia de los Estados Unidos de Norteamérica se encuentra entre los partidarios de las pruebas y de la ley NCLB, y tiene previsto solicitar al Congreso 2 700 millones de dólares en fondos adicionales para la educación pública (más del doble de lo que se pidió para el año fiscal 2014).

La ley NCLB promueve la rendición de cuentas en las escuelas públicas y sus defensores argumentan que ha contribuido a mejorar el aprendizaje entre las minorías, estudiantes de bajos recursos y los que tienen una competencia limitada de inglés.

El Secretario de Educación de los Estados Unidos, Arne Duncan se pronunció por una reforma a la ley de la era de rendición de cuentas del ex presidente Bush, la NCLB, y a pesar de que se ha incrementado el número de detractores de las evaluaciones estandarizadas, Duncan señaló que las pruebas deberán permanecer como obligatorias porque los padres de familia y los maestros tienen el derecho y la necesidad de conocer el progreso anual de los estudiantes. Duncan recomienda que el tiempo que se invierte en la preparación para tomar los exámenes no sea excesivo ni distraiga el tiempo escolar para la enseñanza, es decir, se busca limitar el tiempo que la escuela invierte en la preparación para los exámenes estandarizados.

<sup>1</sup> <http://www.eldiariony.com/examenes-obligatorios-matematicas-lectura-medio-lucha-politica> 12/01/2015  
[http://www.huffingtonpost.com/2015/01/12/no-child-left-behind-2015\\_n\\_6453092.html](http://www.huffingtonpost.com/2015/01/12/no-child-left-behind-2015_n_6453092.html) 12/01/2015

### 3.1. INTRODUCCIÓN

En Inglaterra, la educación obligatoria se divide en cuatro etapas clave (*key stages*): etapa clave 1, de los 5 a los 7 años; etapa clave 2, de los 7 a los 11 años; etapa clave 3, de los 11 a los 14 años; y etapa clave 4, de los 14 a los 16 años. Las dos primeras etapas constituyen la educación primaria, y las etapas 3 y 4, la educación secundaria. La educación secundaria comprende hasta los 18 años de edad, pero su obligatoriedad sólo es hasta los 16 años; el periodo no obligatorio de la educación secundaria se cataloga como etapa clave 5.

El currículo nacional en Inglaterra acaba de renovarse; la implementación del nuevo currículo inició en septiembre del 2014 (Department for Education, 2014a). Algunas de las decisiones que se tomaron incluyeron: que las escuelas usen sus propios métodos para evaluar y monitorear el desempeño de los alumnos; las evaluaciones nacionales se mantienen, y buscan ser más “rigurosas y respetadas” (Department for Education, 2014a: 4), y la remoción del reporte de logro y progreso de los alumnos a partir de niveles. Las evaluaciones del ciclo escolar 2013-2014 y 2014-2015 aún tendrán como referencia la valoración por niveles. A partir del ciclo escolar 2015-2016, se realizarán las modificaciones acordadas en la reforma.

La Agencia de Estándares y Pruebas (STA, por sus siglas en inglés) es la responsable de desarrollar e implementar las pruebas para evaluar a los alumnos en tres diferentes periodos de la educación básica: a) educación preescolar (*early years*), b) etapa clave 1; y c) etapa clave 2 (STA, 2014a). La etapa clave 1 termina en el segundo año de la educación primaria, cuando los alumnos tienen aproximadamente 7 años de edad. La etapa clave 2 termina en el sexto año de educación primaria, cuando los alumnos tienen aproximadamente 11 años.<sup>1</sup> Los exámenes aplicados en estos periodos son conocidos como Standard Attainment Tests (SAT). Su grado de formalidad va incrementándose, es decir, la aplicación de las pruebas de la etapa clave 1 es menos formal, por ejemplo, el periodo de aplicación abarca varias semanas, y los docentes también tienen la posibilidad de elegir el tipo de tareas a las que se someterá a los alumnos. A partir de la etapa clave 2, se incrementa la formalidad de las aplicaciones, se especifican ciertos días de aplicación, y el procedimiento es también más controlado.

Al final de las etapas clave 4 y 5, se aplican también pruebas a los estudiantes. En la etapa 4, los estudiantes presentan un grupo de exámenes que se denominan Certificado General de Educación Secundaria (GCSE, por sus siglas en inglés), regularmente presentan exámenes sobre 8 o 10 asignaturas, y al finalizar la etapa clave 5 presentan los exámenes del Certificado General de Educación de Nivel Avanzado (Advanced Level Exams).

<sup>1</sup> Con respecto a la etapa clave 3, conviene señalar que en el 2008 se decidió que no se aplicaran evaluaciones al final de este periodo.

Para el desarrollo de una prueba, la documentación señala que se requiere un periodo de tres años (Standards & Testing Agency, 2014e). De hecho, para las pruebas que se aplicarán en el 2016, ya se ha elaborado un marco de referencia que utilizan los desarrolladores. Las pruebas son administradas por el personal de las escuelas.

Para la revisión que se hizo, se consideraron las pruebas de etapa clave 2. Cabe señalar que el término de diversidad no se especifica dentro de la documentación de las pruebas; sin embargo, especificaciones sobre la justicia de las evaluaciones, minimización de sesgos y maximización de accesibilidad fueron utilizadas como ejes para la identificación de las prácticas que se presentan en este apartado.

## 3.2. DESCRIPCIÓN GENERAL DE LA EVALUACIÓN O EVALUACIONES

### a) Propósito de la evaluación

Las pruebas de la etapa clave 2 comenzaron a ser aplicadas en 1995. En los marcos para el desarrollo de éstas se señala que su propósito principal es “determinar el logro alcanzado por los alumnos con relación a los objetivos de aprendizaje señalados en el currículum” (Standards & Testing Agency, 2014: 4).

### b) Objeto de la evaluación

El objeto de las evaluaciones de la etapa clave 2 es el logro educativo de los estudiantes. En el 2014, las evaluaciones externas (*statutory assessments*) comprendieron las siguientes áreas: lectura, gramática, puntuación, ortografía (*spelling*) y matemáticas (Department for Education, 2014c: 2). Además, se aplica una prueba de ciencias a una muestra de estudiantes, y los docentes emiten una evaluación de cada uno de los alumnos al finalizar el ciclo escolar con respecto a las siguientes asignaturas: inglés, lectura, escritura, matemáticas y ciencia.

### c) Alcance: subsistema, escuelas, alumnos

Las pruebas son aplicadas a todos los alumnos de la etapa en cuestión en Inglaterra. En Gales las evaluaciones las realiza el docente, para las mismas etapas clave que el sistema educativo inglés. En Escocia, se aplican evaluaciones muestrales anualmente.

En la aplicación de 2014, 532 102 alumnos presentaron las pruebas de etapa clave 2 de matemáticas.

A partir de 2014 la prueba de ciencias se aplica, cada dos años, a una muestra de 9 500 niños seleccionados aleatoriamente en 1 900 escuelas (se eligen 5 niños por escuela) (Standards & Testing Agency, 2014a: 6). La aplicación de esta prueba se hace después de las pruebas censales; por ejemplo, en el 2014, la aplicación de las pruebas censales fue en la semana del 12 al 15 de mayo, la prueba censal se comenzó a aplicar el 2 de junio, y el control de la aplicación fue realizado por personal externo (Standards & Testing Agency, 2014a: 6).

## d) Población a la que se dirige: niveles y grados

Los exámenes de etapa clave 2 están dirigidos a alumnos con una edad aproximada de 11 años. Se dirige a todos los alumnos de Inglaterra. La tabla 11 muestra los diferentes tipos de escuelas, y la participación que se espera que tengan en las pruebas Key Stage 2:

**Tabla 11** Tipos de escuelas en Inglaterra y la participación que tienen en las pruebas Key Stage 2

Tipo de escuela	Descripción	Relación con aplicación de pruebas Key Stage 2
Escuelas subvencionadas ( <i>maintained schools</i> )	Escuelas financiadas por las autoridades locales. Están obligadas a seguir el currículum nacional.	La evaluación y el reporte de las acomodaciones aplica para todas las escuelas subvencionadas, incluyendo las especiales.
Academias en Inglaterra ( <i>Academies</i> )	Escuelas independientes financiadas por el Estado. No están obligadas a seguir el currículum nacional.	Deben aplicar las evaluaciones en las etapas clave, cuando el acuerdo de financiamiento así lo establece.
Unidades de remisión o canalización alumnos ( <i>pupil referral units</i> ), escuelas subvencionadas en hospitales, y niños educados en casa	<p>Las unidades de remisión de alumnos son establecimientos que atienden a alumnos que no pueden asistir a las escuelas ordinarias (niños con enfermedades temporales o permanentes, con problemas de comportamiento, con fobia a la escuela; madres adolescentes, adolescentes embarazadas o niños sin lugar en escuela regular).</p> <p>Las escuelas en hospitales admiten sólo a niños cuyo tratamiento médico requiere hospitalización.</p> <p>Los padres de familia pueden decidir educar a sus hijos en casa (Government Digital Service, 2014). Si éste es el caso, tienen que notificar su decisión a la escuela en la que se encuentren registrados y a las autoridades educativas locales. En estos casos, las autoridades locales registran a los alumnos, y emplean a un consultor en educación en casa para que apoye a las familias (City of York Council, 2015).</p>	<p>En los dos primeros casos, los alumnos están obligados a tomar los exámenes del currículo nacional.</p> <p>Los niños que son educados en casa no presentan los exámenes, a menos que se encuentren dentro de las listas de alguna escuela que participe en la aplicación.</p>
Escuelas en el extranjero	<p>Comprenden escuelas que atienden a hijos de personal que se encuentra en las fuerzas armadas, ministros de defensa, y organizaciones financiadas por los ministros de defensa que se encuentran en el extranjero.</p> <p>También puede haber otras escuelas británicas en el extranjero.</p>	<p>Las escuelas que atienden a los hijos de personal en servicio de defensa deben participar en estas evaluaciones.</p> <p>Otras escuelas no pueden participar formalmente en estas evaluaciones, pero pueden solicitarlas para su aplicación local.</p>
Escuelas de educación especial no subvencionadas	Escuelas de educación especial aprobadas por la Secretaría de Estado que no son mantenidas por el Estado. Cobran colegiaturas, pero son instituciones sin fines de lucro (Douglas Silas Solicitors, 2015).	No están obligadas a participar, aunque pueden hacerlo si lo desean.
Escuelas independientes	Escuelas privadas que pueden tener fines de lucro, no están obligadas a seguir el currículum nacional, pero que deben estar registradas y ser inspeccionadas con regularidad. (Government Digital Service, 2015)	No están obligadas a participar, aunque pueden hacerlo si lo desean.

Fuente: elaboración propia a partir de Standards & Testing Agency (2014a: 11-12) o las fuentes citadas en el interior de la tabla.

## e) Impacto: usos y consecuencias pretendidos de los resultados

Los usos que se espera se hagan de los resultados de esta evaluación son:

- Hacer responsables a las escuelas por el aprendizaje y progreso de sus alumnos.
- Informar a los padres y escuelas secundarias sobre el desempeño de los niños en lo individual.
- Permitir las comparaciones entre escuelas, así como monitorear el desempeño a nivel local y nacional (Standards & Testing Agency, 2014: 4).

## 3.3. PRÁCTICAS DE ATENCIÓN A LA DIVERSIDAD

### 3.3.1. Conceptualización de la diversidad

El marco regulatorio de las pruebas, destinado a proteger los intereses de los alumnos, señala la expectativa de que todos los alumnos tengan acceso a las pruebas y que éstas tengan las siguientes características: adecuación a su edad; medición efectiva de sus habilidades, destrezas y desarrollo de conceptos; equidad para todos los alumnos sin importar su "género, lengua, religión, origen étnico y social, o discapacidad".

La accesibilidad de las Evaluaciones Nacionales (*National Assessments*) a las cuales pertenece Key Stage 2, ha sido una preocupación desde su origen (Office of Qualifications and Examinations Regulation, 2011a: 6), tratando de que las preguntas sean accesibles al mayor rango de estudiantes posible.

Aunque el término de diversidad no se encuentra presente en la documentación analizada, asumimos que las características de los alumnos a las cuales se refiere cuando se enuncia la preocupación por la justicia de las evaluaciones tienen que ver con la atención a la diversidad.

### 3.3.2. Tipos de diversidad que considera

Las medidas implementadas en cuanto al diseño de las pruebas y las condiciones de aplicación están relacionadas con los siguientes grupos de alumnos (Standards & Testing Agency, 2014a, 2014b):

#### a) Niños con NEE:

- Debilidad visual
- Limitaciones auditivas
- Lesiones o enfermedades
- Dificultades para concentrarse
- Dificultades para entender diagramas bidimensionales

#### b) Niños con dominio limitado de la lengua inglesa

Es importante señalar que, aun cuando un niño no sea clasificado dentro de las categorías antes señaladas, pueden solicitar las acomodaciones disponibles de acuerdo con sus necesidades.<sup>2</sup>

En la documentación referente a la evaluación de la accesibilidad de las pruebas, se hacen alusión a otros grupos en los que interesa asegurar que no haya sesgo en los ítems: niños que padezcan dislexia o dificultades de aprendizaje específicas; varones pertenecientes a la clase trabajadora inglesa (*white English working-class boys*); niños que provienen de las comunidades migrantes (*children from traveller community*) (Office of Qualifications and Examinations Regulation, 2011a: 7-8).

### 3.3.3. Atención a la diversidad en las diferentes fases del proceso evaluativo

#### 3.3.3.1. En el diseño y aplicación de los instrumentos de evaluación

En el Marco Regulatorio para las Evaluaciones Nacionales se establecen los criterios con base en los cuales se juzgará el cumplimiento de los acuerdos en cada una de las evaluaciones. Los criterios comunes se muestran en la Tabla 12:

**Tabla 12** Criterios comunes para juzgar el cumplimiento de los acuerdos de las evaluaciones

<b>Validez</b>
La evaluación debe generar resultados que provean una medida válida del conocimiento, las habilidades y el entendimiento que los estudiantes requieren demostrar tal como están especificados en los objetivos de la evaluación. Validez es un concepto central en la evaluación de la calidad de las evaluaciones. Como consecuencia, los organismos que participan en la evaluación deberán contar con procesos y procedimientos que aseguren y generen la evidencia para apoyar la interpretación y uso de los resultados. La validez es el primer criterio de revisión y los demás criterios se encuentran incluidos en éste. Por esa razón, la validez será el objeto de las actividades de revisión.
<b>Confiabilidad</b>
La evaluación generará resultados que provean una medida confiable del desempeño de los alumnos.
<b>Comparabilidad</b>
La evaluación generará resultados que sean comparables a través del tiempo.
<b>Minimización de sesgo</b>
La evaluación deberá minimizar el sesgo, diferenciado solamente a los estudiantes con base en sus habilidades para cumplir los requisitos del Currículo Nacional y los aprendizajes de la primera infancia.
<b>Manejabilidad</b>
La evaluación deberá ser factible de llevarse a cabo; las exigencias hacia las escuelas y los alumnos deberán ser razonables, de tal manera que el proceso de evaluación esté balanceado con la utilidad de los resultados. <sup>3</sup>

Fuente: Traducción propia a partir de Office of Qualifications and Examinations Regulation (2011b: 16, 29).

Como parte del proceso de diseño de las pruebas, se pide que los organismos a cargo de este proceso elaboren una especificación de la prueba que incluya los siguientes elementos: descripción de la forma de la evaluación, cómo se espera que alcance sus propósitos y cumpla

<sup>2</sup> Por ejemplo, para los niños que padecen dislexia, de acuerdo con sus necesidades individuales y las prácticas a las que están habituados en las aulas, pueden solicitar alguna de las siguientes acomodaciones: tiempo adicional para responder la prueba, lector, escriba, uso de computadora, software de apoyo, impresión de los cuadernillos en papel de color, descansos durante la aplicación, entre otras (British Dyslexia Association, 2015).

<sup>3</sup> Se especifica por ejemplo que se espera que las fechas estipuladas para el cumplimiento de ciertas actividades por las autoridades locales y las escuelas sean claras, alcanzables y publicadas oportunamente, que no representen una carga para los actores participantes.

los criterios comunes; la cobertura de la evaluación con respecto a los programas de estudio y la manera en la cual los organismos responsables cumplirán con sus funciones (Office of Qualifications and Examinations Regulation, 2011b: 16)

De acuerdo con la Agencia de Estándares y Pruebas (STA, por sus siglas en inglés) el proceso de diseño de las evaluaciones sigue principios de Diseño Universal. En los marcos de desarrollo de cada una de las pruebas se especifica que deben cumplir con las siguientes condiciones (Standards & Testing Agency, 2014), como se muestra en la tabla 13:

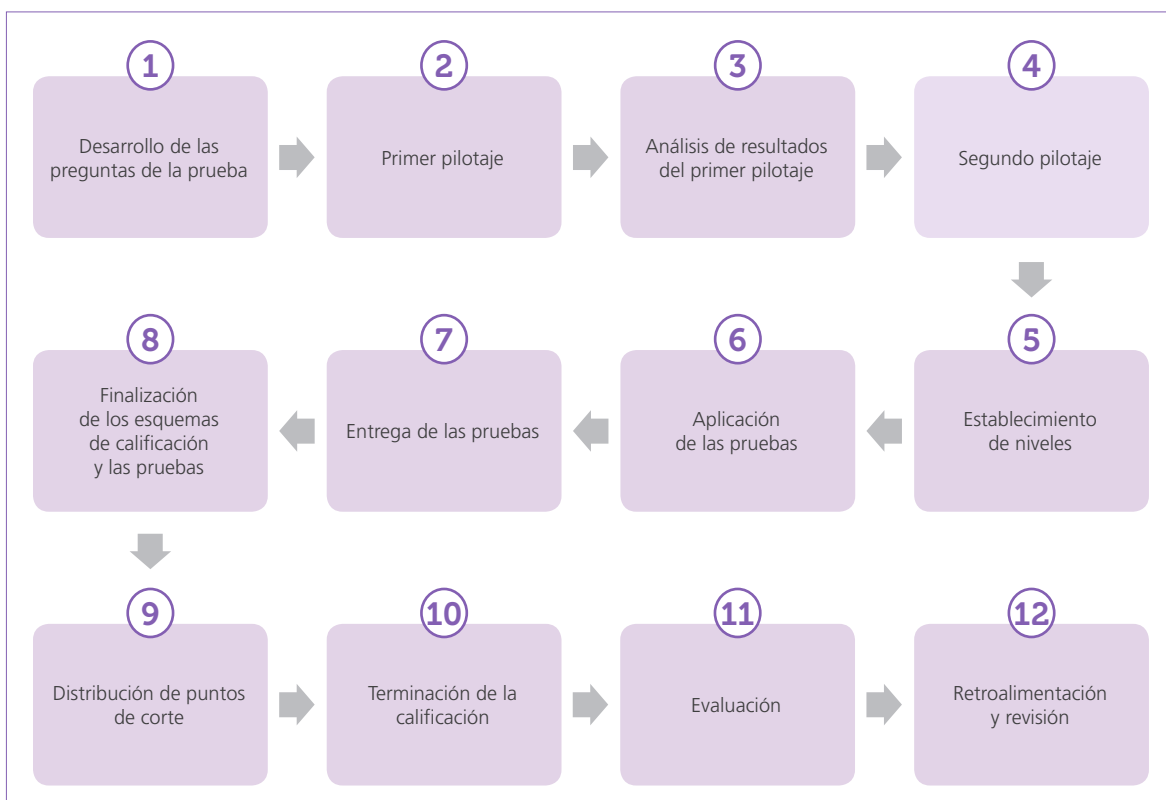
**Tabla 13** Condiciones que deben cumplir las Evaluaciones Nacionales

- Usar medios adecuados para que los alumnos demuestren sus conocimientos, habilidades y destrezas
- Proveer desafíos adecuados para todos los niños y generar oportunidades para que puedan alcanzar estándares altos
- Proveer oportunidades a todos los niños para mostrar su desempeño, sin importar su género, NEE, discapacidad, contexto social, cultural y lingüístico
- Usar materiales familiares para los niños que participen en la prueba y para los que estén adecuadamente preparados para usarlos
- No afectar la autoestima y confianza de los niños
- Estar libres de estereotipos y todo tipo de formas de discriminación

Fuente: adaptado de Standards & Testing Agency (2014: 42).

La figura 2 ilustra el proceso general que se sigue en las Evaluaciones Nacionales:

**Figura 2** Proceso general de las Evaluaciones Nacionales



Fuente: elaboración propia a partir de QCDA (2010).

En la descripción del proceso se señala que, desde la elaboración de la primera versión de los ítems, se incluyen consideraciones sobre su accesibilidad. Las agencias responsables del diseño de las evaluaciones ponen a prueba sus ideas iniciales sobre las preguntas que podrían ser incluidas. En estas pruebas se involucra un pequeño número de escuelas, buscando que estén representadas las escuelas de educación especial, escuelas que cuentan con alumnos con NEE, y que haya niños para quienes el inglés no es su primera lengua.

El proceso de desarrollo de las pruebas se lleva a cabo en un periodo de dos a tres años. Antes de la aplicación definitiva se realizan dos pilotajes. Antes del primer pilotaje se conforman paneles de expertos y docentes que revisan y dan retroalimentación sobre la adecuación de los ítems en términos de contenido y contexto.

El primer pilotaje persigue los siguientes propósitos: “determinar cómo responden los niños a cada pregunta; garantizar que puedan entender la redacción de cada pregunta y no haya confusiones; asegurar que las imágenes sean adecuadas y constituyan un apoyo; obtener datos confiables sobre la dificultad de cada pregunta” (QCDA, 2010: 6). La muestra de escuelas que participan en esta aplicación es diseñada por una agencia distinta a los desarrolladores de las pruebas. En la muestra se busca que haya una distribución geográfica amplia de las escuelas; que se incluyan escuelas de diferentes tamaños y tipos (*v. gr.*, escuelas de educación especial y las que cuentan con alumnos con NEE). La aplicación también la realizan externos a la agencia desarrolladora, en este caso, la misma agencia contrata consultores que coordinen la aplicación de la prueba.

El segundo pilotaje se realiza cuando los ítems se han terminado de elaborar y se han conformado los cuadernillos. Se busca obtener datos sobre el desempeño de la prueba para contrastarla con respecto a la aplicación anterior y poder llevar a cabo la equiparación. Como resultado de este pilotaje se esperan cambios menores.

Entre el primer y el segundo pilotajes, se elaboran las pruebas modificadas para niños con limitaciones visuales y auditivas. El desarrollo de estas pruebas lo hace una agencia especializada en modificación de pruebas (QCDA, 2010: 11), quien también se encarga de elaborar las orientaciones para su aplicación y calificación, así como del entrenamiento que deberán recibir los calificadores.

Es importante señalar que el ciclo de desarrollo de las pruebas ha incluido desde el inicio de las Evaluaciones Nacionales una revisión exhaustiva de todas las preguntas, por expertos que han trabajado con alumnos de los diferentes grupos en los que busca garantizar la accesibilidad de la prueba, particularmente los niños con NEE (Office of Qualifications and Examinations Regulation, 2011a: 6).

Con respecto a las pruebas de Key Stage 2 en inglés, la Fundación Nacional para la Investigación Educativa (NFER, por sus siglas en inglés) ha señalado que el desarrollo de las evaluaciones incluye poner exhaustivamente a prueba los materiales con muestras representativas de niños; a través de estas pruebas se valora el funcionamiento técnico de los ítems, lo mismo que su adecuación. También se incluyen paneles de docentes, otros profesionales educativos y expertos, para valorar la adecuación y accesibilidad para todos los niños a los que se les aplicará la prueba (National Foundation for Educational Research, 2015).



Además de las orientaciones sobre los procesos que realizan los organismos que participan en el desarrollo de las evaluaciones, se promueve el monitoreo y la evaluación de los procesos de la evaluación. La Oficina Regulatoria de Calificaciones y Examinaciones recupera evidencias que permiten valorar la calidad del proceso, e implementa revisiones antes de la aplicación de las pruebas, para asegurar que cumplan con sus propósitos y los criterios comunes establecidos (Office of Qualifications and Examinations Regulation, 2011b: 19).

A continuación se describen las acomodaciones que se hacen para la atención de niños con necesidades de evaluación diversas.

### 3.3.3.2. Atención de niños con Necesidades Educativas Especiales

La atención de niños con NEE es un apartado detallado dentro de los manuales de aplicación. El tratamiento de las NEE es complejo, y requiere la colaboración entre docentes, directores y padres de familia en cuanto a la toma de decisiones sobre las adecuaciones o acomodaciones necesarias para facilitar el acceso de los alumnos a la prueba.

Las NEE se entienden como las dificultades para el aprendizaje que son experimentadas por algunos niños. Estas dificultades pueden ocurrir cuando: *a)* los niños muestran una dificultad mayor en aprender que la mayoría de los niños de su misma edad; *b)* tienen una discapacidad que les impide o limita el uso de los recursos educativos que generalmente son provistos a los niños de su misma edad en las escuelas de la autoridad educativa local; *c)* se encuentran en edad de escolaridad obligatoria y caen dentro de las definiciones *a* o *b* o se encontrarían dentro de ellas si una provisión de educación especial no se les hubiera otorgado" (Department for Education and Skills, 2001: 6).

Es importante señalar que se sugiere que las diferencias entre la lengua materna que hablan los alumnos y la lengua en que se realiza la enseñanza en la escuela no sean consideradas como características que apunten a dificultades de aprendizaje (Department for Education and Skills, 2001: 6). Esto coincide con la postura de Solano-Flores (2012a), quien señala que un dominio limitado de una segunda lengua no puede ser considerado como una deficiencia; cuando se asume como una deficiencia, puede sentar las bases para una clasificación equivocada de estos alumnos dentro del grupo de niños con discapacidad u otras NEE.

Las discapacidades se definen como: "una limitación física o mental que tiene un efecto adverso en el largo plazo en su habilidad [del alumno] para realizar actividades cotidianas" (Department for Education and Skills, 2001: 7). También se definen algunas discapacidades específicas, tales como: ceguera, sordera y mudéz.

Uno de los puntos de partida importantes con respecto a la atención la población estudiantil que tiene dificultades aprendizaje es que se establecen estándares de desempeño y descriptores específicos que obligatoriamente deben utilizarse para reportar su progreso y sus logros (Department for Education, 2014b). Los docentes eligen las actividades de las pruebas que se presentan a los alumnos con base en el nivel de logro en el que se ubican los estudiantes. Ciertamente, a partir de la implementación del nuevo currículum, estas definiciones cambiarán, dado que los niveles de logro desaparecen como una de las medidas para reportar el progreso de los alumnos.

En la documentación que guía la administración de las pruebas se señala que en el diseño y la aplicación de las pruebas las agencias participantes deberán asegurar que se han dado las provisiones necesarias para que niños con NEE participen en la aplicación.

Algunas de las NEE que se atienden en estos procesos de evaluación son las que se describen enseguida.

#### Niños con debilidad visual

Para la atención de niños con debilidad visual, se cuenta por lo general con los siguientes recursos impresos: versiones en braille de la prueba; versión modificada de la prueba que implica que los diagramas sean suprimidos o modificados, el tamaño de letra ampliado y la prueba impresa en una hoja de mayor tamaño (Modified Large Print o MLP); e impresión ampliada, que consiste en la misma prueba, pero en una hoja de mayor tamaño y con un tipo de letra también mayor (Enlarged Print o EP). Cuando las escuelas requieren usar este tipo de recursos no necesitan autorización de las autoridades educativas locales, pero sí se requiere solicitar estas versiones con anterioridad.

Por otro lado, también se da la opción de que se hagan adaptaciones a los folios de la prueba. Los cuadernillos de la prueba se abren una hora antes de que comience la aplicación, y se pueden hacer algunas adaptaciones, siempre y cuando éstas representen un elemento común a las prácticas empleadas en las aulas. Algunas de las adaptaciones son: ampliar o reducir los folios de los cuadernillos o secciones de ellos, enriquecer diagramas, hacer adaptaciones en el grado de braille de la prueba, utilizar versiones grabadas, y el uso de dispositivos técnicos o electrónicos como ayuda para los sustentantes.

Con respecto a la modificación de los folios, se proveen orientaciones específicas que señalan los casos en los cuales no se podrán hacer cambios, y se dispone de una línea telefónica de atención especial para consultar las modificaciones que se pueden realizar.

El enriquecimiento de diagramas puede comprender: recalcar el sombreado de diagramas y gráficas para aumentar la claridad visual; adicionar líneas para aquellos niños que tengan dificultades de percepción espacial; los diagramas se pueden ampliar, cortar y montar sobre tarjetas u otros materiales, cuidando de no cambiar la naturaleza de las preguntas; fotocopiar en papel de colores, usar láminas de colores o lentes con filtros de colores; añadir énfasis en las pruebas de impresión modificada y braille. Cuando los diseñadores de la pruebas consideran que algún diagrama no puede ser modificado, se señala de manera específica y se pide a los docentes y al director ser riguroso en el seguimiento de estas orientaciones, de tal manera que no se dé una ventaja a estos alumnos para responder la prueba (STA, 2014b, 2014c, 2014d).

Para los niños que regularmente tienen acceso a grabaciones o a personas que les leen, se pueden utilizar versiones grabadas de algunas pruebas, en especial, de matemáticas. Las instrucciones se pueden clarificar con las versiones grabadas, pero no se pueden hacer cambios a las frases ni a su orden. No se permiten versiones grabadas de las pruebas de lengua inglesa.

En cuanto a los dispositivos técnicos o electrónicos, se promueve el uso de aquellos que sean regularmente utilizados por niños con visión baja, tales como circuito cerrado de televisión y digitalizadores de texto. Se pide a los docentes que reporten el uso de estos dispositivos en

un formato en donde se especifican: los datos de la escuela y el niño, la información sobre la prueba en la cual se usaron estos dispositivos, la razón de su uso, y si fue utilizado durante toda la prueba o parcialmente. El formato debe contener la firma del director, donde confirma que el alumno no fue ayudado en la materia evaluada, y que el lenguaje, puntuación y fraseado es propio del alumno.

Para los niños que utilizan las versiones en braille y que recurren al apoyo de un lector, éste puede ayudar durante la aplicación, mediante el uso de diagramas, gráficas y tablas táctiles, para obtener información que la prueba estandarizada proporciona a niños que no son ciegos.

### Niños con limitaciones auditivas

Se encontraron precisiones para facilitar el acceso a la prueba de matemáticas mentales que a continuación se describen.

Los niños que tienen limitaciones auditivas usan una hoja de respuestas sin tiempo adicional y materiales de estímulo. En las pruebas de matemáticas se puede usar la grabación provista en un disco compacto, o el aplicador puede leer la versión estándar de la prueba individualmente o a pequeños grupos de niños.

Se dan también diferentes opciones para la administración de la prueba, tales como: administrarla individualmente con el volumen más alto que sea posible, usar audífonos o hacer lectura en voz alta al niño. Cuando se elige esta última opción, se notifica a través de un formato, previa solicitud dirigida a las autoridades.

Se puede recurrir también a otros recursos para facilitar el acceso de niños con limitaciones auditivas a la prueba, tales como: subtítulo, uso de materiales modificados, y modificaciones de acceso específicas.

El subtítulo puede utilizarse si la escuela no usa el lenguaje de señas británico, y si ésta es una práctica habitual en la escuela. Los subtítulos se generan a través de la computadora. Cuando se necesita el uso de subtítulos, se requiere hacer una solicitud expresa para abrir anticipadamente las pruebas, de tal manera que tengan tiempo suficiente para generarlos.

Los materiales modificados se proveen para niños con pérdida auditiva permanente o de largo plazo que dependen del lenguaje de señas británico u otra comunicación a través de señas, o para quienes complementan su audición restante con lectura de labios. Estos materiales comprenden: guión modificado para el administrador de la prueba; un disco compacto o tarjetas con los materiales estímulo de la hoja de repuestas, además de materiales de estímulos adicionales que puedan apoyar el acceso al contenido de la prueba. Para el uso de estos materiales se proveen indicaciones específicas, incluyendo el espacio de la aplicación. Por ejemplo, la lectura del guión por el especialista en lenguaje de señas debe realizarse en un salón por separado, sea de uno a uno, o en pequeños grupos.

A los niños que tengan una limitación auditiva profunda y que no utilizan un sistema de señas o lectura de labios para comunicarse, se les puede administrar la prueba usando subtítulos en los cuales cada pregunta se repite dos veces en la pantalla, y se designa un tiempo específico para que la respondan.

### Lesiones o enfermedades

Cuando los niños sufren alguna lesión durante una semana previa a la aplicación, la cual afecta su habilidad para contestar el examen, por ejemplo, con una fractura de mano o brazo, se permite utilizar algunas de las siguientes adecuaciones: 25% de tiempo adicional, un escriba, un procesador de palabras o una transcripción de texto hablado. Aunque en estos casos no se requiere autorización, sí deben notificar la modificación elegida, especialmente cuando hayan requerido un escriba y transcripciones.

Cuando la lesión haya ocurrido con más de una semana de anticipación para la aplicación, se debe dar a los niños la oportunidad de practicar con un escriba en condiciones similares a las de la prueba. Si esto ocurre, no será elegible para solicitud de tiempo adicional durante la aplicación.

Con respecto a enfermedades, se precisa que no se espera que los niños que se encuentren enfermos el día de la prueba la sustenten. Para estas situaciones, se deben registrar en las listas como ausencias. Si hay una enfermedad menor y el niño no puede presentar la prueba más tarde durante el día, se puede considerar hacer una variación en el calendario de la aplicación; estas variaciones pueden ser hasta 5 días posteriores a la fecha oficial.

### Niños con dificultades para concentrarse

Para los niños que tienen problemas de atención severos se sugiere el uso de apuntadores, siempre que éstos formen parte de las actividades cotidianas en el aula. Los apuntadores deben ser los mismos asistentes para el aprendizaje que han sido asignados a los niños. Para su apoyo en la aplicación, se requiere que se pongan de acuerdo anticipadamente sobre la mejor manera de ayudar a los niños. El apoyo se brinda de manera individual y su principal fin es atraer la atención de los niños nuevamente a la tarea que están realizando en la prueba. Los apuntadores deben tener especial cuidado en no realizar acciones que puedan ser interpretadas como una ayuda adicional.

Los recordatorios que dan a los niños pueden ser verbales. Si un niño encuentra dificultades para concentrarse en ciertas preguntas, los apuntadores también pueden usar notas adhesivas para cubrir otras preguntas en la página. Estas prácticas sólo pueden realizarse cuando son utilizadas como un apoyo regular a los niños. No requieren solicitar permiso a las autoridades locales, pero el uso de apuntadores debe estar justificado.

#### 3.3.3.3. Características de recursos de apoyo específicos

En este apartado se detallan las características de algunos apoyos específicos que facilitan el acceso de los alumnos a la prueba. En primer lugar se muestran adecuaciones con respecto al tiempo, después con respecto a personal de apoyo, y finalmente se hace alusión a los recursos tecnológicos y otros recursos materiales que pueden permitirse en las aplicaciones.

## Modificaciones en cuanto al uso del tiempo

- Tiempo adicional

Se puede utilizar tiempo adicional cuando los niños requieren comúnmente de una mayor cantidad de tiempo para realizar actividades en las aulas; también se considera que el tiempo adicional puede ser conveniente cuando los alumnos se encuentran trabajando justo en el nivel evaluado por los exámenes.

El uso de tiempo adicional para el examen requiere autorización previa. Por ejemplo, durante el ciclo escolar 2014-2015, se aplicarán estas pruebas en mayo de 2015, y se ha dado como límite hasta el 14 de abril para que las escuelas realicen la solicitud de tiempo adicional (Standards & Testing Agency, 2014c).

A los alumnos que han sido declarados con NEE y quienes usan las versiones en braille o las impresiones ampliadas, se les concede automáticamente tiempo adicional. A los primeros se les concede hasta 25% de tiempo adicional para todas las pruebas, excepto las de matemáticas mentales y ortografía. A los segundos se les puede conceder hasta 100% de tiempo adicional.

En la documentación también se especifica a qué alumnos no se les deberá asignar tiempo adicional, dentro de los que se encuentran: los que están por debajo del nivel de las pruebas y en el nivel de las pruebas pero sin poder participar, aun con ajustes en el acceso (Standards & Testing Agency, 2014b).

- Apertura temprana de las pruebas

Se puede solicitar permiso para abrir los cuadernillos más de una hora antes del inicio reglamentario de la prueba. Este tipo de permisos puede solicitarse cuando se requiere hacer adaptaciones al texto o preparar herramientas o equipo para los alumnos que lo necesitarán. Algunos de los casos son: cuando se requiere preparar equipo y materiales táctiles para niños que utilizan las versiones modificadas de la prueba impresa (MLP) y en braille; cuando una parte o la totalidad de la prueba necesita ser ampliada para tener un tamaño de letra mayor, ya sea para niños con debilidad visual o con NEE; cuando los lectores o intérpretes de lenguaje de señas necesitan coordinar la presentación de una prueba a un gran número de niños con debilidad auditiva o que usan lenguaje de señas; cuando los niños usan equipo o adaptaciones de texto como parte de las prácticas cotidianas en aula y esto no puede ser preparado en la hora anterior a que comience la prueba; cuando el niño se encuentra registrado en programas educativos dirigidos a niños con NEE o que tenga una declaración de NEE.

Cuando se autoriza, el director funge como el responsable de salvaguardar la seguridad y confidencialidad de los materiales de la prueba antes de su aplicación y hasta que se envíen para calificación. El personal de la escuela debe esperar a recibir la autorización para abrir los materiales, y sólo el director o la persona que haga las adaptaciones tendrán acceso a las pruebas una vez abiertas.

No se permite la apertura anticipada de las pruebas en los siguientes casos: cuando un niño está trabajando por debajo del nivel a evaluar y no obtendrá beneficios al sustentar

la prueba incluso con modificaciones de acceso, cuando el niño no usa equipo especial o adaptaciones a los textos como parte de las prácticas cotidianas en la escuela, cuando las necesidades de los niños pueden ser atendidas a través de otras modificaciones como descansos o trabajar en un cuarto separado.

Cuando se utilizan versiones modificadas de la prueba, se puede obtener autorización para abrirlas hasta dos días antes de la fecha establecida. En estos casos, se muestra en la portada de los paquetes la fecha en que pueden ser abiertos los paquetes con materiales para la prueba; cada versión modificada de las pruebas está acompañada de guía por asignatura para su aplicación; las orientaciones para su ensamblaje y los contenidos son confidenciales.

- Descansos durante la aplicación

Los periodos de descanso durante la aplicación pueden ser útiles cuando un niño enfrenta dificultades para concentrarse o experimenta fatiga. Los descansos pueden hacerse dividiendo la prueba, o parando el reloj.

La división de la prueba se hace regularmente durante la hora previa a la aplicación. En estos casos, los docentes deben asegurarse de que se administren todas las secciones en el día programado, que los niños tengan el mismo tiempo previsto para completar la prueba que quienes lo hacen sin descansos, y que se mantenga el orden de las preguntas.

Si se para el reloj, será importante asegurarse que no se dé tiempo adicional sin autorización.

Durante los descansos no puede haber discusiones sobre el contenido de la prueba. El uso de estos periodos no requiere autorización especial. Se dan indicaciones específicas para el uso de este apoyo en algunas pruebas. Por ejemplo, para las pruebas de lectura de los niveles 3 a 5, se señala que pueden darse descansos una vez que el alumno haya leído un texto particular o respondido las preguntas de ese texto (antes de cambiar a otra prueba); en el caso del nivel 6, se pueden dar descansos durante la lectura de un texto, pero se debe asegurar que el niño haya leído completamente el texto antes de comenzar con las preguntas.

### Incorporación de personal de apoyo en la aplicación

- Lectores

Los lectores se emplean durante la aplicación sólo cuando forman parte de una práctica cotidiana en el salón de clases. Son normalmente los mismos maestros o asistentes de apoyo, no pueden ser otros alumnos, familiares de los niños o tutores. Los lectores no necesitan ser especialistas en la materia evaluada, pero deben ser capaces de leer con precisión y a una velocidad adecuada; el apoyo que brindan lo deben dar de forma individualizada a los niños. Antes de la prueba es necesario que conozcan su formato y su estilo, el papel que desempeñarán, y que identifiquen lo que sí pueden y lo que no pueden leer. La lectura incluye la posibilidad de leer las propias respuestas de los alumnos si ellos lo requieren.

El apoyo de los lectores por lo general es necesario cuando la edad de lectura de los niños es menor a su edad actual. No se requiere apoyo de lectores cuando los niños son capaces de leer el material por sí mismos, y no se necesita pedir permiso a autoridad local.

En las pruebas de inglés (lengua, referido a comprensión de la lectura), sólo se pueden leer las instrucciones generales (lo que está registrado en la portada e indicaciones que no sean parte de las preguntas). En estas pruebas no se pueden leer los textos, preguntas o partes de las respuestas de los niños.

En la prueba de gramática, puntuación y ortografía, sólo se permiten los lectores si son una práctica cotidiana del salón de clases. Se provee un documento con indicaciones específicas sobre qué leer y qué no leer en cada pregunta (Standards & Testing Agency, 2014d).

En las pruebas de matemáticas, los lectores pueden clarificar las instrucciones, pero no pueden: dar información adicional, clarificar el significado de vocabulario específico de la asignatura; referir a los alumnos a partes previas de la pregunta o partes de las preguntas. Se especifica que sólo pueden leer palabras y números pero no los símbolos matemáticos.

- **Escribas**

Un escriba es un asistente que registra las respuestas dictadas directamente por los alumnos. Antes de considerar el uso del escriba, se pide a los maestros que valoren utilizar un procesador de palabras o transcriptor. Si estas opciones no se consideran adecuadas y el uso del escriba es una práctica cotidiana en las aulas para el niño en cuestión, se puede utilizar sin autorización previa aunque se requiere la notificación.

Los escribas se recomiendan principalmente cuando los niños, debido a sus características físicas, no son capaces de escribir por sí mismos las respuestas a través de un procesador de textos; cuando el niño puede escribir pero tiene una limitación de movimiento que le causa malestar o incomodidad al escribir; cuando escriben lentamente o con mucha dificultad, o cuando no pueden escribir debido a una lesión.

Los escribas también pueden utilizarse durante una parte de la prueba en las siguientes circunstancias: cuando se sabe que los niños experimentan fatiga, y cuando se requiere hacer más claro el texto escrito por los alumnos, sea por que cuentan con limitaciones visuales o motoras que los hacen escribir sobre su propio texto. En estos casos, los niños pueden comenzar la prueba en el horario programado y utilizar el escriba cuando sea necesario.

Los escribas deberán cumplir con las siguientes condiciones: ser capaces de escribir de manera legible, escribir a una velocidad razonable, y tener conocimiento práctico sobre la tarea a realizar; asimismo pueden también fungir como lectores. Por otra parte, los escribas no pueden ser familiares de los niños o tutores.

No se permite que los escribas transcriban braille, pero sí pueden hacer correcciones a un escrito cuando un niño lo solicite. Los escribas deben trabajar de acuerdo con el ritmo de los niños, sin apresurarlos, y seguir de manera precisa las instrucciones de los niños para dibujar o añadir diagramas o gráficas en las pruebas de matemáticas.

## Otras modificaciones para facilitar el acceso a la prueba

- Asignación de puntos compensatorios

Los puntos compensatorios pueden ser otorgados a niños que tienen una profunda deficiencia auditiva y que no usan lenguaje de señas o lectura de labios. Sólo pueden ser asignados en ciertas pruebas, por ejemplo, en los niveles 3 a 5, en las pruebas de ortografía y matemáticas mentales; en el nivel 6, en las pruebas de gramática inglesa y ortografía. Los puntos compensatorios se basan en el puntaje promedio generado durante el pretest en el desarrollo de la prueba. Otorgar estos puntos puede hacer que los niños sean asignados a un nivel específico, lo que dependerá de que hayan completado los otros componentes de la prueba y del puntaje total. Si un niño toma la prueba, aprobada para consideración especial, se le otorgarán puntos basados en su desempeño en la prueba. No se le asignarán puntos compensatorios.

- Traducciones escritas u orales

Se hace uso de traducciones escritas u orales sólo cuando forman parte de prácticas cotidianas en aula. No se requiere autorización y el director es el responsable de asegurar la calidad de la traducción.

En los exámenes de lengua inglesa no se permiten traducciones orales o escritas; tampoco se puede ayudar con la lectura y comprensión de las preguntas o pasajes del texto en los que éstas se basan. Sólo se pueden traducir las instrucciones generales en la portada y las indicaciones que no son parte de las preguntas.

En los exámenes de matemáticas, los niños pueden leer las preguntas en inglés y responderlas también en esa misma lengua, o leerlas en inglés y responderlas en su primera lengua. En este último caso, se debe realizar una transcripción utilizando al traductor con el que normalmente trabaja el alumno; el documento original del alumno también debe ser enviado para el marcado, y se debe realizar una notificación.

En los casos en que varios niños requieran traductores, se puede considerar hacer modificaciones en la programación de la aplicación.

- Transcripciones de los escritos de los alumnos.

No se necesita autorización para crear transcripciones, pero debe llenarse un formato de notificación. Se dan orientaciones específicas para crear una transcripción. La transcripción puede hacerse cuando es muy difícil leer el manuscrito de los niños. Si la persona que calificará el examen puede leer el manuscrito, entonces se califica la versión original.

Las transcripciones sólo pueden hacerse al finalizar un examen; se puede utilizar un color diferente cuando se hacen transcripciones parciales sobre el manuscrito de los alumnos. Las transcripciones completas deben hacerse en folios nuevos.

Las transcripciones se hacen en presencia de los niños, antes de que dejen el aula; durante este tiempo el alumno del cual se hace la transcripción deberá mantenerse separado del resto del grupo, hasta que la transcripción sea terminada. Se pide que se asegure



que las respuestas originales no sean cambiadas, y no se da ayuda con ortografía, puntuación y fraseado. A los calificadores se les envía la versión original y la transcripción.

- Procesadores de palabras y otras ayudas técnicas o electrónicas.

Al igual que en el caso de las transcripciones, no se necesita solicitar autorización, pero sí es necesario reportar la utilización de estos implementos. Se pueden utilizar aquellos que normalmente requieren los alumnos en las actividades de aula. Dentro de estos implementos se pueden encontrar: amplificadores ópticos o electrónicos, lectores digitales, computadoras, o registradores de notas en braille usados para procesadores de texto.

Para el uso de estos recursos se pide que no haya ningún contenido del alumno o de la materia grabado dentro del equipo, se deben suspender las funciones de autocorrección y revisión de ortografía, y no pueden tener acceso a internet o intranet.

#### 3.3.3.4. Consideraciones adicionales para facilitar el acceso a las pruebas

Es necesario señalar que también se consideran modificaciones específicas para niños que tienen las siguientes características:

- Dificultades para entender diagramas bidimensionales. Cuando éste es el caso, se pueden proporcionar objetos reales similares a los que se incluyan en las pruebas. Las personas que administren la prueba deberán señalar en el papel cuándo han utilizado este tipo de objetos de apoyo.
- Niños con doble discapacidad sensorial. Se contempla la utilización de varios recursos de apoyo. Se dan guías específicas sobre recursos complementarios y excluyentes.
- Niños con fluidez limitada en lengua inglesa. Se puede optar por una traducción oral de la prueba, siempre y cuando la versión traducida se comunique oralmente, los tiempos se respeten y la traducción no sea hecha por un familiar del niño.

### 3.4. OPINIÓN PÚBLICA

Se revisó la prensa disponible vía electrónica: *The Guardian*, *The Telegraph* y *BBC Noticias*. En diciembre de 2014 se publicaron los resultados de las evaluaciones nacionales, y en enero de 2015 se presentó un sistema electrónico desarrollado para que los usuarios de los resultados hicieran sus propias comparaciones con respecto a criterios específicos.

En la prensa de diciembre, se encontraron noticias relativas a las tendencias en el desempeño de los estudiantes; aunque no tenían que ver con la accesibilidad de las pruebas, sí aludían a los resultados de grupos de estudiantes específicos, tales como: estudiantes para quienes el inglés no es su primera lengua y quienes provienen de contextos socioeconómicos desfavorecidos. En general, en los resultados del 2014 se observan mejores desempeños en estos grupos de alumnos. La prensa resalta la reducción de la brecha de desempeño de acuerdo con el contexto socioeconómico de los alumnos, y los resultados destacados de los alumnos para quienes el inglés no es la primera lengua en regiones específicas (*i.e.* centro de Londres). Prevalecen comparaciones entre tipos de escuelas y regiones geográficas.

### 3.5. ESTUDIOS EVALUATIVOS SOBRE LAS EVALUACIONES NACIONALES

En el 2012 se llevó a cabo una revisión independiente de las pruebas Key Stage 2 (Bew, 2011). Algunas de las recomendaciones relacionadas con la justicia de las evaluaciones y la atención a la diversidad fueron: generar un indicador del progreso de los alumnos —esto es especialmente importante para los alumnos que obtienen un bajo desempeño en las evaluaciones—; con relación a los estudiantes con necesidades educativas especiales, se solicita que sus logros sean reconocidos y celebrados en el sistema de rendición de cuentas; para los estudiantes con capacidades sobresalientes, se enfatiza la necesidad de continuar aplicando pruebas que sean adecuadas y representen un desafío para sus capacidades.

En el 2014 se publicó un estudio que tuvo como objetivo comparar el desempeño de diferentes subgrupos de estudiantes en la muestra de Ciencias de 2010 y 2011 (He, Anwyll, Glanville y Opposs, 2014). Los subgrupos de alumnos estuvieron clasificados por género, grupo étnico, inglés como una lengua adicional, elegibilidad para comidas gratuitas en la escuela, y NEE. Se analizó el funcionamiento diferencial de los ítems en lo individual (DIF) y de la prueba en conjunto (funcionamiento diferencial de la prueba, DTF por sus siglas en inglés). Los resultados muestran que las pruebas de ciencias analizadas funcionan de manera similar entre los diferentes grupos tanto a nivel de ítems en lo individual como en la prueba completa.

Es importante señalar que en 2011 se emitieron orientaciones para monitorear el acceso a las evaluaciones nacionales del currículo; en éstas se plantearon dos propuestas: conformación de paneles de docentes para la revisión de la accesibilidad de los reactivos, y estudios sobre funcionamiento diferencial de los ítems (Office of Qualifications and Examinations Regulation, 2011a).

#### 4.1. DESCRIPCIÓN DEL SISTEMA EDUCATIVO

Australia tiene un sistema escolar federal que incluye seis estados y dos territorios y la responsabilidad de la educación recae en ellos. Las escuelas pueden ser públicas o privadas. Las escuelas públicas (*government*) operan de manera directa bajo las directrices del ministerio de educación, ya sea estatal o territorial, mientras que las escuelas privadas (*non-government*) operan bajo las condiciones establecidas por las autoridades de los estados y los territorios, y pueden ser parte de un sistema de escuelas o ser completamente independientes. La mayoría de las escuelas privadas tienen una afiliación religiosa, en un porcentaje superior con la iglesia católica, y es así que genéricamente la educación privada se conoce como “católica” e “independiente” (Santiago, Donaldson, Herman y Shewbridge, 2011).

En Australia la educación está conformada por un periodo obligatorio de 12 años de escolarización dividido en un nivel preescolar, con duración de un año, seguido del nivel primaria que dura de 6 a 7 años, y la educación secundaria que dura de 5 a 6 años ya sea que se curse o no el 7º año de primaria, dependiendo del estado o territorio. Una vez concluido el periodo obligatorio, los alumnos pueden ingresar a la enseñanza superior que puede ser de tipo universitario o vocacional.

Los territorios y los estados han sido responsables del establecimiento del currículo y de los estándares de logro, y en 2008 los ministerios de educación acordaron el desarrollo y la implementación de un currículo nacional de los grados 1 a 12 para reforzar las áreas de inglés, matemáticas, ciencias e historia a partir de 2011, con la expectativa de avanzar en las otras áreas del currículo en los años subsecuentes.

Otra preocupación del sistema educativo australiano es cerrar la brecha en los resultados de logro de los estudiantes aborígenes e isleños del Estrecho de Torres y los otros estudiantes australianos, mediante un plan de acción que apoya las metas del Plan Nacional de Aborígenes, la Política Educativa Isleña del Estrecho de Torres y la Declaración de Melbourne sobre los Objetivos de la Educación de Jóvenes Australianos.

El Plan de Acción identifica las acciones nacionales, sistémicas y locales en seis ámbitos de prioridad estratégica que pueden contribuir a mejorar los resultados en la educación de los aborígenes y los isleños del Estrecho de Torres, incluyendo:

- momento en el que el estudiante está listo para asistir a la escuela
- compromiso y conexiones
- asistencia
- alfabetización y aritmética

- liderazgo, calidad de la enseñanza y desarrollo del personal
- opciones reales después de la escuela

El marco de responsabilidad y rendición de cuentas, establecido a través del Plan de Acción, atribuye los ámbitos de responsabilidad de los objetivos, indicadores de resultados y acciones. Incluye también un reporte público anual y una estrategia de evaluación continua para asegurar que las acciones después de 2014 estén fundamentadas e informadas de la mejor evidencia de la que se disponga (Santiago *et al.*, 2011).

#### 4.1.1. Australian Qualifications Framework (AQF)

El Marco Australiano de Reconocimientos o Certificaciones (AQF, por sus siglas en inglés), distingue al sistema educativo australiano; se trata de una política nacional que inició en 1995 y concede los reconocimientos o certificaciones de la educación secundaria y terciaria, así como de la educación superior y la formación profesional. Esta política nacional incorpora las calificaciones o reconocimientos de cada sector educativo y formativo en un único marco nacional integral de calificaciones.

En este contexto, la educación y la formación o capacitación son una responsabilidad compartida entre los gobiernos estatales y territoriales: los ministerios de educación, de formación y de empleo son responsables de la AQF; a la par, el sector industrial, los sindicatos, las asociaciones profesionales y las autoridades gubernamentales tienen interés en la calidad de la formación de los estudiantes.

La AQF establece las normas y estándares para emitir los títulos o certificaciones australianas, y está conformada en una estructura que rige a las instancias autorizadas para otorgar los reconocimientos.

El marco organizativo de la AQF es una estructura taxonómica de los niveles y tipos de reconocimientos y de resultados de aprendizaje ordenados jerárquicamente. Este enfoque está diseñado para permitir la consistencia en la forma en que las calificaciones se describen, así como la claridad acerca de las diferencias y relaciones entre los tipos de calificaciones.

Uno de los objetivos fundamentales del AQF es facilitar los elementos para otorgar calificaciones o certificaciones formales y vincular las calificaciones del AQF entre sí de distintas maneras: entre las escuelas, la formación profesional y la educación superior, y la educación y el trabajo de acuerdo a los intereses de los estudiantes; es decir, contribuye a la planificación de las carreras, ya que le permite al estudiante transitar de un nivel de estudio a otro con facilidad, y de una institución a otra.

El AQF establece itinerarios de aprendizaje flexibles, transparentes y sistemáticos, y la eliminación de las fronteras entre los sectores educativos, y está orientado a preparar al estudiante para continuar estudiando o para su inserción en el mercado laboral (Australian Qualifications Framework Council, 2013).

## 4.2. SISTEMA DE EVALUACIÓN EDUCATIVO

Australia tiene un sistema exhaustivo de evaluación que opera en cuatro niveles principales:

1. nacional y sistémico (estado, territorio o sistema no gubernamental);
2. la escuela, a través de una variedad de formas de evaluación escolar, normalmente en el contexto de un marco de mejora del rendimiento de la escuela;
3. el maestro, a través de registros, desempeño y habilidades avanzadas de enseñanza; y
4. el estudiante, a través de instrumentos que van desde las pruebas estandarizadas nacionales, a la evaluación formativa diaria en la escuela.

El marco general de evaluación aparece como altamente sofisticado y bien conceptualizado, especialmente en su nivel superior (niveles nacional y sistémico).

Las características particularmente positivas del marco de evaluación incluyen: las metas educativas nacionales como un punto de referencia, la fuerte capacidad a nivel nacional para dirigir la evaluación, un enfoque en los resultados de los estudiantes, un sistema coherente de evaluación para el aprendizaje, una buena estructura para integrar la rendición de cuentas y la mejora, y un compromiso con la transparencia en la comunicación de los resultados. Las prioridades para el desarrollo evaluativo incluyen: el establecimiento de estrategias nacionales para el fortalecimiento de los vínculos con la práctica en el aula, promover una mayor consistencia nacional al tiempo que se da cabida a la diversidad local, la integración del sector educativo no gubernamental en el marco general de evaluación, el desarrollo de vínculos dentro del marco general, y el mantenimiento de los esfuerzos para mejorar la capacidad de evaluación y valoración (Santiago *et al.*, 2011).

### 4.2.1. Programa Anual de Evaluación Nacional en Alfabetización y Aritmética (NAPLAN)

La Autoridad Australiana de Currículo, Evaluación y Reporte (ACARA, por sus siglas en inglés) implementa el Programa de Evaluación Nacional en Alfabetización y Aritmética (NAPLAN, por sus siglas en inglés). Este programa se aplica a estudiantes de 3º, 5º, 7º y 9º de sectores público y privado, y evalúa habilidades en lectura, escritura, ortografía, gramática, puntuación y aritmética. NAPLAN es considerado como un marco consistente para evaluar el aprendizaje de los estudiantes y sus resultados gozan de amplia credibilidad entre la comunidad escolar (ACARA, 2014a, 2014b, 2014c, 2014d).

El NAPLAN está enmarcado en el Programa Nacional de Evaluación (NAP, por sus siglas en inglés), que incluye pruebas trienales de alfabetización en tecnologías de la información y la comunicación (6º y 10º), alfabetización científica (6º), y educación cívica y ciudadana (6º y 10º). El NAPLAN incluye también la participación en el Programa para la Evaluación Internacional de Estudiantes (PISA, por sus siglas en inglés) y en el Estudio de Tendencias Internacionales en Matemáticas y Ciencia (TIMSS, por sus siglas en inglés), lo que permite el monitoreo del progreso de los estudiantes tomando como referencia los estándares internacionales.

De acuerdo a sus desarrolladores, el NAPLAN refleja los contenidos y habilidades que por lo general se encontraban en las distintas evaluaciones de los estados y territorios administradas antes del NAPLAN. Las pruebas se desarrollan en colaboración con los estados y territorios, el gobierno de Australia, y los representantes de las iglesias católicas y los sectores escolares independientes.

Cada centro escolar recibe los resultados individualizados de cada estudiante, además de un análisis del conjunto de los estudiantes, por franjas y en niveles. El reporte individual del alumno es enviado al padre o tutor. En el reporte se señala, para cada habilidad medida, el promedio, el estándar nacional mínimo y la franja en que cae 60% de los alumnos en su grado. Se indica de igual manera si el alumno alcanzó o no el estándar nacional mínimo. El reporte incluye una tabla con una breve descripción de las habilidades típicamente demostradas por alumnos en cada nivel (Santiago *et al.*, 2011).

### 4.3. PRÁCTICAS DE CONSIDERACIÓN DE LA DIVERSIDAD

La Declaración de Melbourne (MCEETYA, 2008) establece que la escolaridad en Australia promueve la equidad y la excelencia en materia educativa, y da a las escuelas la responsabilidad de reducir los efectos de la desventaja educativa asociada al estatus socioeconómico, la condición indígena y la discapacidad. En este contexto, la equidad es definida en dos dimensiones: horizontal, cuando los servicios son igualmente accesibles a todos los miembros de la comunidad con niveles similares de necesidades; y vertical, cuando los servicios representan las necesidades especiales de determinados grupos de la comunidad y distintos aspectos de la prestación de servicios se ajustan para satisfacer estas necesidades. La condición de “necesidades especiales” puede incluir factores geográficos, culturales o de otro tipo que impidan el acceso a un servicio estándar (Santiago *et al.*, 2011).

En el proceso de desarrollo del NAPLAN no está claro si ha recibido suficiente atención la accesibilidad para los estudiantes indígenas y los estudiantes con discapacidades. La tensión entre las pruebas estandarizadas, que se suponen comunes a todos los estudiantes, y la necesidad de mostrar sensibilidad al conocimiento de la cultura local y de la experiencia es difícil de resolver.

Los análisis técnicos de los resultados del NAPLAN 2008 sugieren que esa tensión no ha sido contrarrestada de manera adecuada para los estudiantes indígenas; es decir, el análisis del funcionamiento diferencial de los ítems sugiere que, en efecto, funcionan de manera diferenciada para los estudiantes indígenas y los no indígenas. De manera particular, en el proceso de desarrollo de la prueba del NAPLAN se consulta a expertos indígenas de los estados y territorios. Ellos proporcionan información concreta sobre la idoneidad de los elementos de la prueba y la idoneidad de los materiales de estímulo para los estudiantes indígenas. En el proceso se realizan exámenes de prueba utilizando una muestra de alumnos, y el análisis de los resultados se lleva a cabo para asegurar que todos los reactivos y preguntas sean culturalmente apropiados y libres de prejuicios. Pese a este esfuerzo por reducir las brechas en el logro educativo, los representantes indígenas están preocupados por aspectos relacionados con la equidad de la prueba a pesar de los esfuerzos del NAPLAN para eliminar el sesgo cultural (Santiago *et al.*, 2011).

En cuanto a la atención a los estudiantes con discapacidades, no obstante que la Declaración de Melbourne alude a la equidad para este tipo de estudiantes, no parece claro si se ha seguido este principio y hasta dónde en el desarrollo del NAPLAN. Existen acomodaciones o servicios especiales que les son permitidos a los alumnos con discapacidad, pero la consistencia con que se administra la prueba es incierta. Además, no se incluyó información de cómo funcionó la prueba para los estudiantes con discapacidad en el informe técnico del proyecto 2008.

Entre las recomendaciones de política educativa derivadas de la evaluación sobresale la encomienda de asegurar que el NAPLAN sea una prueba justa para todos los subgrupos. De entre éstos, la accesibilidad para los estudiantes indígenas y estudiantes con discapacidades también merece atención. La evidencia disponible sugiere desafíos en torno al funcionamiento de la prueba para los estudiantes indígenas, mientras que los datos sobre la prueba para estudiantes con discapacidades están ausentes. Los desarrolladores de la prueba tienen el reto de considerar en su diseño y desarrollo estrategias para mitigar las deficiencias identificadas y reunir pruebas para documentar el éxito en esta área, y para garantizar que el NAPLAN sea una prueba justa para los estudiantes indígenas. Es preciso continuar involucrando a los representantes indígenas en la especificación de la prueba inicial y en el proceso de desarrollo de los ítems para coadyuvar a la sensibilidad cultural de la prueba y disminuir la percepción de que la prueba muestra sesgos. De acuerdo a los desarrolladores de las pruebas, estas acciones deben combinarse con una buena comunicación de los esfuerzos del NAPLAN con los representantes indígenas para que los resultados sean confiables en la óptica de la población indígena.

De la misma manera, las acomodaciones o servicios especiales para los estudiantes con discapacidades necesitan ser reforzados para reflejar de manera sistemática el tipo de apoyo y asistencia que los estudiantes reciben en el aula. No obstante que existen acomodaciones disponibles, la consistencia con que son administradas es todavía una cuestión abierta y el conocimiento de la validez de los datos está ausente en gran medida. El objetivo es hacer del NAPLAN una prueba integradora basada en el principio de que todos los estudiantes tienen la oportunidad de participar en actividades educativas, incluyendo las actividades de evaluación, para demostrar sus conocimientos, habilidades y competencias.

Los servicios especiales o acomodaciones disponibles para los estudiantes que tienen discapacidades pero que son capaces de completar toda o partes de las pruebas, se enlistan a continuación. Estos ajustes deben reflejar el tipo de apoyo al que el estudiante accede de manera regular en el aula. Los padres, cuidadores y el estudiante mismo participan en las decisiones sobre estos ajustes a la discapacidad.

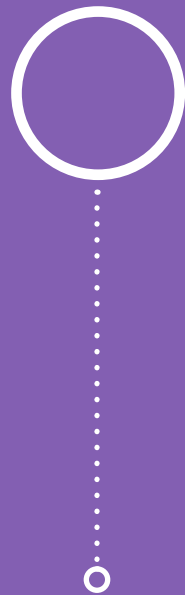
1. Impresión de los documentos de evaluación en formato grande o en braille.
2. Impresión en blanco y negro o copias en papel de color.
3. Lector de pantalla a color o superposiciones u otro uso de tecnología.
4. Computadora para la tarea de escribir con la comprobación de ortografía y gramática apagada.
5. Una persona de apoyo para la aritmética, pruebas de lectura y lenguaje para rellenar los alvéolos de las respuestas y para escribir en las preguntas de respuesta corta.
6. Un escribano especialmente entrenado para la escritura cuando ésta es la rutina habitual de apoyo en el aula.
7. Tiempo extra para la prueba (e. g., aproximadamente cinco minutos por cada media hora).

8. Descansos durante la prueba (e. g., aproximadamente cinco minutos por cada media hora).
9. Apoyos orales o de señalética para los casos de deficiencia auditiva.
10. PDF interactivos para los estudiantes que no pueden utilizar lápiz y papel.

#### 4.4. OPINIÓN PÚBLICA

No se encontraron documentos relevantes que den cuenta del tratamiento de las evaluaciones en la prensa australiana.





# Prácticas internacionales de atención a la diversidad

## Organismos internacionales

## 5.1. DESCRIPCIÓN GENERAL Y PROPÓSITO

El Programa Internacional de Evaluación de los Alumnos (PISA, por sus siglas en inglés) es un proyecto de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) que evalúa de manera permanente y trienal la formación de los estudiantes de 15 años, cerca del final de la educación obligatoria (OECD, s. f.). PISA busca medir los conocimientos y habilidades que son esenciales para la plena participación de los individuos en las sociedades modernas. La evaluación no sólo determina si los estudiantes pueden reproducir los conocimientos; también examina cómo los estudiantes pueden extrapolar lo que han aprendido y aplicar ese conocimiento en entornos desconocidos, tanto dentro, como fuera de la escuela. Este enfoque refleja el hecho de que las economías modernas recompensan a los individuos no por lo que saben, sino por lo que pueden hacer con lo que saben.

El propósito central es conocer el nivel que han adquirido los estudiantes de habilidades necesarias para participar plenamente en la sociedad, centrándose en tres dominios claves: Lectura, Ciencias y Matemáticas. Para PISA, estos dominios están definidos como competencias (*literacies*) científica, lectora y matemática (*El Programa PISA de la OCDE, ¿Qué es y para qué sirve?, s/f*).

La medición de conocimientos y habilidades de los estudiantes de 15 años de edad ofrece información sobre cómo los países y las economías participantes están desarrollando a los futuros talentos. De acuerdo a la encuesta para medir las habilidades de los adultos del Programa para la Evaluación Internacional de las Competencias de Adultos (PIAAC, por sus siglas en inglés), existe una estrecha correlación entre el desempeño de los países en los diferentes ciclos de PISA y la competencia de los grupos de edad correspondientes en el nivel de comprensión lectora y la capacidad de cálculo de los países de la OCDE (OECD, 2013).<sup>1</sup>

Los estudios de PISA se llevan a cabo en un gran número de países, lo que implica que se trata de una evaluación que comprende diferentes culturas y sistemas escolares, y que la prueba se tiene que aplicar en distintos idiomas, es decir, mediante baterías de pruebas traducidas y adaptadas en cada lengua de los países participantes. A las áreas principales que evalúa PISA,

<sup>1</sup> La OCDE ha puesto en marcha un nuevo proyecto para valorar el nivel de conocimientos y el grado y distribución de destrezas de la población adulta a lo largo de su vida laboral, entre los 16 y los 65 años. Esta evaluación se encuadra en un marco comparativo coherente para poder establecer contrastes internacionales.

El PIAAC mide las destrezas cognitivas básicas que permiten a las personas adultas participar en la vida social y económica del siglo XXI, y también las habilidades laborales básicas que les exige su puesto de trabajo. Para ello, se recoge información sobre los antecedentes personales de la población seleccionada en la muestra, además de información sobre las exigencias del puesto de trabajo, el uso de tecnologías de la información y la comunicación y, en general, el nivel de comprensión lectora (*literacy*) y la capacidad de cálculo (*numeracy*).

competencias lectora, matemática y científica, a partir de 2012, se agrega una cuarta, la resolución de problemas (OECD, 2014).

Los resultados de PISA ofrecen información para la política y la práctica educativas, y ayudan a monitorear las tendencias en la adquisición de conocimientos y habilidades de los estudiantes en distintos países y en diferentes subgrupos demográficos dentro de cada país. Los resultados permiten conocer y medir los conocimientos y habilidades de los estudiantes en sus propios países en comparación con los de otros y establecer objetivos de política en relación con las metas medibles a través de decisiones sobre el alcance y naturaleza de la evaluación.

## 5.2. MARCO CONCEPTUAL

El desarrollo de los marcos conceptuales del proyecto OCDE/PISA se ha llevado a cabo de forma continua desde que se creó el programa en 1997, y consiste en una sucesión de pasos que abarca:

- La definición de trabajo para el área de conocimiento junto con la descripción de los supuestos en que se basa dicha definición.
- La valoración del modo de organizar los ejercicios para poder informar a los investigadores y a los encargados de las políticas educativas sobre el rendimiento de los alumnos en cada una de las áreas de conocimiento evaluadas, junto con la identificación de las características clave que deben tenerse en cuenta al elaborar los ejercicios de evaluación destinados a un uso internacional.
- La determinación de las características básicas a las cuales se ajustará la elaboración de las pruebas, con definiciones basadas en la bibliografía existente y en la experiencia en la realización de otras evaluaciones a gran escala.
- La validación de estas variables y la valoración de la contribución de cada una de ellas a la dificultad del ejercicio de la prueba en los diferentes países participantes.
- La preparación de un esquema de interpretación de los resultados (OECD, 2006; Rico, 2006).

La elaboración y la validación de un marco conceptual para cada una de las áreas de conocimiento redundan en una mejor medición; además, aportan otras ventajas potenciales:

- Un marco conceptual proporciona un vehículo y un lenguaje comunes para debatir el propósito de la evaluación y lo que se pretende medir. Este debate fomenta el desarrollo de un consenso sobre el marco conceptual y los objetivos de la medición.
- Un análisis de las destrezas y de los tipos de conocimiento asociados a un rendimiento positivo aporta la base necesaria para establecer estándares o niveles de rendimiento. Conforme avanza la comprensión de lo que se está evaluando y la capacidad para interpretar los resultados sobre una escala concreta, se puede desarrollar una base empírica para transmitir un conjunto más rico de información a los diferentes tipos de público a los que se dirige el proyecto.
- La identificación y comprensión de las variables concretas que constituyen la base de un rendimiento positivo incrementa la capacidad para evaluar lo que se está midiendo y realizar cambios en la evaluación a lo largo del tiempo.

- La comprensión de lo que se está midiendo y su conexión con lo que decimos sobre los estudiantes proporciona un importante vínculo entre la política educativa pública, la evaluación y la investigación, lo cual, a su vez, refuerza la utilidad de los datos recogidos (OECD, 2006).

### 5.3. MARCO CONTEXTUAL

Además del examen, los estudiantes contestan un cuestionario de contexto socioeconómico. Dado que se utilizan muestras representativas de cada país participante, se pueden realizar inferencias del país en su totalidad, pero no por regiones o estados.

Con objeto de recabar información contextual, PISA pide a los alumnos y a los directores de los centros que respondan a un cuestionario de una duración aproximada de 30 minutos. Estos cuestionarios desempeñan un papel crucial, pues permiten analizar los resultados en función de las diversas características de los alumnos y de los centros de enseñanza. A través de los cuestionarios se pretende obtener información sobre los alumnos y su entorno familiar, incluyendo su capital económico, social y cultural; diversos aspectos de la vida de los alumnos, como por ejemplo, su actitud hacia el aprendizaje, sus hábitos y su vida en el entorno escolar y familiar; diversos aspectos sobre los centros de enseñanza, como por ejemplo, la calidad de los recursos humanos y materiales de los centros, el carácter público o privado de su gestión y financiación, los procesos de toma de decisiones y las prácticas del personal docente; el contexto del aprendizaje, incluyendo el tipo y las estructuras de las instituciones, el tamaño de las clases y el grado de implicación de los padres; las estrategias de aprendizaje autorregulado; las preferencias motivacionales y la orientación hacia la consecución de los propios fines; los mecanismos cognitivos del yo, las estrategias de control de la acción, las preferencias por determinados tipos de situaciones de aprendizaje, los métodos de aprendizaje y las habilidades sociales necesarias para un aprendizaje de tipo cooperativo o competitivo; así como diversos aspectos del aprendizaje y la instrucción, entre otros aspectos (OECD, 2006).

### 5.4. CARACTERÍSTICAS DE LA PRUEBA

El diseño y traducción de la prueba, así como el muestreo y la recolección de datos, están sujetos a estrictos controles de calidad, lo que da como resultado que PISA se considere altamente válida y confiable (OECD, 2014, s. f.). Las características de PISA se enlistan a continuación:

- Tiene una orientación política que vincula resultados con datos de antecedentes y actitudes de los estudiantes hacia el aprendizaje y con los factores claves que dan forma a su aprendizaje dentro y fuera de la escuela, con el fin de resaltar las diferencias de rendimiento e identificar las características de los estudiantes, las escuelas y sistemas escolares que muestran un buen desempeño.
- Maneja un innovador concepto de *alfabetización*, que se refiere a la capacidad de los alumnos de aplicar conocimientos y habilidades en materias claves y de analizar, razonar y comunicar eficazmente cómo identificar, interpretar y resolver problemas en una variedad de situaciones.

- Recoge información de los estudiantes sobre su motivación para aprender, sus creencias acerca de sí mismos y sus estrategias de aprendizaje.
- Sigue una administración regular, lo que permite a los países y a las economías monitorear su progreso en el cumplimiento de sus objetivos de aprendizaje clave.
- Muestra una creciente ampliación de su cobertura: PISA 2012, abarcó los 34 países miembros de la OCDE y 31 economías y países socios.

## 5.5. COBERTURA

La población de estudiantes participantes es definida por normas técnicas estrictas. Se encontró necesario reducir la cobertura de la población objetivo a través de la exclusión, por ejemplo, de una región geográfica remota y pequeña que, por razones de inaccesibilidad, dificultaba la aplicación, o de un grupo por razones de lengua, o por razones políticas, organizativas u operativas, o bien, por necesidades de educación especial de los estudiantes. En un estudio de esta naturaleza es importante definir de manera constante los tipos de exclusión para todos los países participantes, y además la tasa de exclusión tiene que ser limitada. Si se excluyera a una proporción significativa de estudiantes o de escuelas, los resultados no serían representativos de una nación, de ahí que se hayan realizado esfuerzos por minimizar las exclusiones de acuerdo a los estándares técnicos de la prueba.

La exclusión puede ser a nivel de escuela o a nivel de estudiantes individuales. Las áreas que son consideradas parte de un país para propósitos de PISA, pero que no resultan incluidas en la evaluación, se consideran como zonas no cubiertas. En este sentido se ha puesto especial atención para asegurar que las exclusiones, de ser necesarias, se minimicen de acuerdo a los estándares técnicos de PISA 2009.

Las reglas de exclusión internacionales dentro de las escuelas se especificaron como sigue:

- Estudiantes discapacitados intelectualmente: son aquellos estudiantes que tienen alguna discapacidad mental o emocional y que en la opinión de personal profesional capacitado tienen un retraso cognitivo que no les permite ser evaluados de manera válida por PISA. Esta categoría incluye a aquellos estudiantes que no pueden seguir las instrucciones generales de la prueba. Los estudiantes no fueron excluidos en términos de bajo rendimiento escolar o por problemas de disciplina.
- Estudiantes con alguna discapacidad funcional: son aquellos que están inhabilitados físicamente de manera permanente, de tal forma que no pueden ser evaluados con validez por PISA. Los estudiantes que están discapacitados funcionalmente y que pueden proporcionar respuestas son incluidos en la prueba.
- Aquellos estudiantes que muestren insuficiencia en el idioma necesitan cumplir los siguientes criterios para ser excluidos:
  - a) no son hablantes nativos de la lengua en que se está evaluando a través de PISA,
  - b) poseen una habilidad limitada en la lengua con que se está evaluando, y
  - c) han recibido menos de un año de instrucción en la lengua de evaluación.

- Estudiantes que no pueden ser evaluados por otras razones, los cuales constituyen un subgrupo que incluye a los individuos con dislexia, disgrafía o discalculia.
- Estudiantes que son instruidos en una lengua para la que no hay materiales disponibles en la evaluación.
- La tasa general de exclusión de un país dado (a nivel escolar e individual) se debe mantener por debajo de 5% de la población objetivo (OECD, 2012: 61).

## 5.6. ATENCIÓN A LA DIVERSIDAD

En el PISA no se identifican medidas específicas de atención a la diversidad, aunque sí se observa una preocupación por asegurar la equivalencia de las pruebas en los distintos países para poder recabar información que sea comparable a nivel internacional (National Project Managers' Meeting, 2010a, 2010b). La equivalencia de las versiones nacionales es un requisito de PISA, lo que significa que la traducción de materiales debe cumplir con estándares rigurosos de calidad en cada una de las versiones nacionales utilizadas por los países participantes; es decir, en las comparaciones internacionales es decisivo contar con lineamientos que garanticen una consistencia mínima en los procedimientos empleados por los diversos países.

Otro componente clave en las comparaciones internacionales es el uso de un sistema de revisión de las traducciones de pruebas. Dicho sistema está basado en el empleo de un equipo de revisión de traducciones en el que no participan los traductores de los países participantes (Solano-Flores, Contreras Niño y Backhoff, 2006: 3). Es crucial asegurar que el proceso de traducción no introduzca sesgos que puedan deformar las comparaciones internacionales; en particular, lo que se busca evitar es:

- hacer la comprensión de los textos, gráficos o tablas utilizadas en las pruebas más fáciles (o difíciles),
- modificar la dificultad de las preguntas para el estudiante de manera no intencional, a través de una formulación que cambia el tipo de estrategia mental requerida,
- introducir ambigüedades que podrían afectar algunas de las variables que se recogen a través de los cuestionarios de fondo (debido a la falta de comparabilidad entre países), o
- adaptar manuales al contexto nacional en formas que cambian los datos recogidos.

Diversos autores señalan que la investigación en el campo de las comparaciones internacionales ha mostrado que la traducción de una palabra, aunque sea ligeramente inexacta, puede ser suficiente para afectar el funcionamiento diferencial de un ítem (Ercikan, 1998, citado en Solano-Flores *et al.*, 2006) y que el uso del procedimiento de *traducción reversa* (basado en volver a traducir el ítem a la lengua original) ya no se acepta como prueba irrefutable de equivalencia entre idiomas. Para asegurar las equivalencias entre las distintas versiones de la prueba, el consorcio responsable de la implementación de PISA hace las siguientes recomendaciones y procedimientos: doble traducción y reconciliación (Solano-Flores, Backhoff y Contreras Niño, 2009).

Para cumplir con las normas de traducción del PISA, es necesario que las versiones nacionales de todos los instrumentos utilizados en la evaluación, la prueba y el cuestionario se desarrollen a través de un procedimiento de doble traducción y reconciliación. Es decir, dos traductores, de manera independiente, traducen el material original en el idioma de destino, y luego una tercera persona debe conciliar estas dos traducciones en una sola versión nacional.

El consorcio desarrolla dos versiones paralelas de la prueba, una en inglés y otra en francés (Sabatini y Bruce, 2011). Se recomienda que los países utilicen la versión en inglés fuente para una de las traducciones en la lengua nacional, y la versión de origen francés para otra. En aquellos países donde encontrar traductores competentes de ambas lenguas fuente sea un problema, una alternativa consiste en doble traducción y reconciliación a partir de uno de los idiomas de origen, seguidos de un extenso cotejo contra la segunda lengua de origen. Como parte de los procedimientos de control de calidad implementados en el PISA, todos los países participantes deberán presentar sus versiones nacionales para verificación contra las versiones inglesa y francesa de origen, por un equipo de verificadores independientes con comandos nativos de los idiomas respectivos del PISA, el cual ha sido especialmente designado y formado para tal efecto.

## Oficina Regional de Educación de la UNESCO para América Latina y el Caribe, Santiago de Chile (OREALC/UNESCO Santiago)

### 6.1. DESCRIPCIÓN GENERAL

La OREALC/UNESCO Santiago tiene como una de sus encomiendas dar seguimiento al cumplimiento de las metas educativas establecidas por los países miembros en el año 2000 en Dakar, iniciativa conocida como *Educación para Todos*. Para realizar su tarea, la OREALC/UNESCO Santiago cuenta con 12 áreas de acción, entre ellas: planificación de la educación, docentes, educación inclusiva, evaluación de la calidad de la educación, e indicadores educativos.

Si bien la Oficina realiza múltiples acciones para promover la atención a la diversidad (Eroles e Hirmas, 2009; Hevia, Hirmas Ready y UNESCO, 2005; Hirmas R. y Carolina, 2008; Varas *et al.*, 2003) en el marco de su área de acción de educación inclusiva, no forman parte de un sistema de evaluación o monitoreo, por lo que se descartaron de este informe y la atención se enfocó en otras áreas.

Las acciones en el área de evaluación de la calidad de la educación se implementan a través del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), organismo que a la fecha ha realizado tres evaluaciones de logro educativo en la mayoría de los países de la región. Los llamados Primer, Segundo y Tercer Estudio Regional Comparativo y Explicativo se realizaron en 1997, 2006 y 2013, respectivamente (Casassus, Froemel, Palafox y Cusato, 1998; LLECE, 2013; Treviño, Castro, Costilla y Acevedo, 2008, 2009), y evaluaron a los alumnos de tercero y sexto grado de primaria en los dominios de lectura, escritura, matemáticas y ciencias (esta última sólo en 6° y a partir del segundo estudio, el SERCE). Pese a que esta evaluación de aprendizajes es una de las más importantes en la región, no se identificaron elementos especialmente destacables de atención a la diversidad (Atorresi *et al.*, 2010; Treviño *et al.*, 2009).

No obstante, el monitoreo de la educación mediante indicadores estadísticos que realiza la OREALC es una experiencia valiosa sobre cómo evaluar el grado en que las personas con algún tipo de discapacidad están incluidas o excluidas de los servicios educativos que ofrece cada país. El Sistema Regional de Información Educativa de los Estudiantes con Discapacidad (SIRIED) constituye un importante esfuerzo de consideración de la diversidad digno de recuperar en este estudio, pues, en el caso mexicano, podría orientar el monitoreo de grupos especiales de estudiantes, desde una perspectiva de inclusión. Esto es así porque aun cuando el SIRIED versa sobre alumnos con discapacidad, su enfoque y metodología bien podrían servir para evaluar, junto con otras herramientas, a subpoblaciones que por sus características específicas no han sido incluidas en las evaluaciones educativas que se realizan en México, tales como niños



migrantes, alumnos cuya lengua predominante es indígena, y estudiantes de educación básica para adultos, entre otras.

El SIRIED nace en 2007 con el propósito general de

contar con un conjunto de información básica regionalmente comparable sobre las necesidades educativas especiales, que permita la construcción de una batería de indicadores significativos con el fin de dar respuesta a los interrogantes centrales que se plantean en la región en relación a esta población y que se constituyan en una herramienta eficaz para formular, analizar y monitorear acciones públicas integrales que garanticen a la región no sólo la igualdad de acceso, sino también la igualdad de oportunidades en la calidad del servicio educativo (SIRNEE, 2007: 10).

El SIRIED se justifica en una fuerte preocupación en la región por asegurar el derecho a una educación de calidad a todas las personas, sin ningún tipo de discriminación, prestando especial atención a los colectivos en situación de mayor vulnerabilidad, lo que es consistente con los planteamientos de la Conferencia Mundial de Educación para Todos de Jomtiem (1990), reafirmados en Dakar (2000). Además,

la carencia de estadísticas e indicadores sobre necesidades educativas especiales es un tema especialmente recurrente. El informe titulado *El derecho a la educación de personas con discapacidades* presentado por el relator especial sobre el derecho a la educación, Vernor Muñoz, destaca con especial énfasis la ausencia de estadísticas sobre la población que presenta discapacidad, así como la proporción de personas con discapacidad que acceden a la escolarización y las características de sus trayectorias escolares (SIRNEE, 2007: 7).

La principal finalidad del SIRIED es contar con un conjunto de información básica e indicadores educativos, regionalmente comparables, que permitan:

- a) Conocer y monitorear la situación educativa de los estudiantes con discapacidad en los países de América Latina en relación con el cumplimiento del derecho a una educación inclusiva y de calidad.
- b) Identificar las barreras que enfrentan las personas con discapacidad así como los recursos y apoyos que requieren para garantizar su pleno acceso a la educación y permanencia, la participación y el aprendizaje.
- c) Proporcionar información relevante para la definición, desarrollo y monitoreo de políticas y la provisión de recursos y apoyos que garanticen la igualdad de oportunidades de la población con discapacidad.
- d) Contribuir al monitoreo de Educación para Todos aportando información sobre una población tradicionalmente excluida de las estadísticas educativas (Blanco, Mascardi y Narvarte, 2010: 11).

En el resto de este apartado, se desarrollan cuatro elementos del SIRIED que ilustran cómo un sistema de información puede contribuir a la conformación de un sistema nacional de evaluación educativa con un enfoque inclusivo: el marco conceptual en que se fundamenta, el proceso que se siguió para su diseño, el modelo de análisis operativo del SIRIED y los primeros resultados obtenidos.

## 6.2. MARCO CONCEPTUAL EN QUE SE FUNDAMENTA

El SIRIED tiene como fundamento cuatro pilares interconectados (Blanco *et al.*, 2010). En primer lugar, el derecho a la educación como un elemento central al cual responde. Se considera la educación como un bien público y derecho humano fundamental del que nadie puede quedar excluido, de lo que se sigue la adopción de un enfoque de derechos en la educación, que ha de basarse en los principios de obligatoriedad y gratuidad, igualdad de oportunidades y no discriminación y el derecho de todos a una educación de calidad a lo largo de la vida.

Su segundo pilar lo constituyen las dimensiones de la calidad de educación desde un enfoque de derechos humanos: relevancia, pertinencia, equidad, eficacia y eficiencia. Estas dimensiones, coincidentes con el modelo del Instituto Nacional para la Evaluación de la Educación (INEE) y que a su vez fueron adoptadas por los ministros de educación de la región de América Latina y el Caribe en Buenos Aires 2007, se analizan desde la especificidad de los estudiantes con discapacidad, considerando algunos elementos que son de especial relevancia para garantizar su acceso, egreso, participación y aprendizaje.

El tercer elemento en que se fundamenta el sistema de información es la educación inclusiva porque es un componente clave del derecho a la educación, de especial significación en el caso de las personas con discapacidad para garantizar su acceso a una educación de calidad en igualdad de condiciones con los demás. Si bien el discurso en educación está claramente orientado hacia la educación inclusiva, las políticas, las normativas y las prácticas todavía se sitúan en la mayoría de los países en el paradigma de la integración, o en un momento de transición hacia la inclusión, lo que hace complejo el establecimiento de indicadores. Los aspectos centrales de ambas posturas se resumen en la tabla 14.

**Tabla 14** Características principales de los paradigmas de integración educativa y educación inclusiva

Integración educativa	Educación inclusiva
El foco de atención son ciertos grupos de estudiantes, tradicionalmente segregados o marginados, que se incorporan a las escuelas normales.	El foco de atención son todos los estudiantes. La finalidad de la inclusión es garantizar una educación de calidad para todos, prestando especial atención a personas excluidas o en riesgo de ser marginados.
La intervención se centra más en la atención individualizada de los estudiantes que en la transformación del sistema educativo y de la enseñanza.	Las acciones están orientadas a transformar la cultura, la organización y las prácticas educativas de las escuelas para responder a las diversas necesidades educativas de todo el alumnado.
Los recursos y sistemas de apoyo están orientados a los alumnos con necesidades educativas especiales (NEE).	Los recursos y sistemas de apoyo están disponibles para todas las escuelas y estudiantes que lo requieran.
La responsabilidad de la educación de los estudiantes integrados recae a menudo en los especialistas o profesionales de apoyo.	Los docentes se responsabilizan del aprendizaje de todos los estudiantes, independientemente de su origen social y cultural y sus características individuales.
Ha supuesto más una transformación de la educación especial que de la educación en general.	Implica una transformación de los sistemas educativos basada en la diversidad y no en la homogeneidad.
En síntesis, bajo este enfoque los estudiantes que se incorporan a las escuelas comunes se tienen que adaptar a la oferta educativa disponible, independientemente de su origen social y cultural, sus capacidades, su lengua o sus situaciones de vida. El sistema educativo permanece inalterable.	En síntesis, bajo este enfoque no se trata sólo que los estudiantes tradicionalmente excluidos se eduquen en las escuelas comunes sino que éstas transformen sus culturas, políticas y prácticas educativas para favorecer su plena participación y aprendizaje. La inclusión busca avanzar más que la integración, transformando la educación en general y el sistema educativo en su conjunto.

Fuente: Blanco *et al.*, 2010.

El cuarto pilar está relacionado con la conceptualización de la población objetivo del SIRIED desde la perspectiva de la atención a la diversidad, una de las características fundamentales de la educación inclusiva. Se parte de la premisa de que las diferencias son consustanciales a la naturaleza humana y que todos los estudiantes tienen unas necesidades educativas propias y específicas para acceder a las experiencias de aprendizaje y participar plenamente en la vida escolar. La educación de los estudiantes con discapacidad se enmarca en el enfoque más amplio de la atención a la diversidad, proporcionándoles los recursos y los apoyos que necesitan, y eliminando las barreras que limitan su acceso y su permanencia en la educación, así como su plena participación y aprendizaje, en coherencia con un enfoque social e interactivo de la discapacidad.

Si bien en sus inicios el Sistema estaba orientado a los estudiantes con necesidades educativas especiales, posteriormente se centró en aquellos que tuvieran alguna discapacidad. El SIRIED asume la concepción de la Convención de los Derechos de las Personas con Discapacidad, que implica un modelo social e interactivo de la discapacidad: “Las personas con discapacidad incluyen a aquellas que tengan deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, al interactuar con distintas barreras, puedan impedir su participación plena y efectiva en igualdad de condiciones con las demás” (ONU, 2006, citado por Blanco *et al.*, 2010: 10).

En esta categorización, la discapacidad se considera un término genérico que incluye déficits, limitaciones en la actividad y restricciones en la participación. Indica los aspectos negativos de la interacción entre un individuo (con una condición de salud) y sus factores contextuales (ambientales y personales). En otras palabras, desde una perspectiva social, la discapacidad se concibe como un fenómeno complejo y multidimensional que surge de la interacción entre factores inherentes a la persona (deficiencia o tipo de discapacidad) y factores del contexto físico y social, cuyas barreras colocan a las personas en situación de desventaja (Blanco *et al.*, 2010). Esto significa que cuanto menores son las barreras y mayores los apoyos, más capaz es la persona de participar en las diferentes áreas de la vida social, aunque siga teniendo una deficiencia.

En correspondencia con una concepción interactiva de la discapacidad, el SIRIED considera datos relativos a los individuos y al contexto educativo. En relación con los sujetos, es necesario cuantificar las personas con discapacidad e identificar sus necesidades en términos de los recursos y apoyos que requieren para acceder a la educación, participar plenamente y aprender. En relación con el contexto educativo, el sistema acopia información mediante 42 indicadores sobre las barreras que enfrentan las personas con discapacidad y sobre la capacidad de respuesta de los sistemas educativos para atender sus necesidades, en términos de apoyos y recursos, y minimizar las barreras que limitan el pleno ejercicio de su derecho a la educación.

El sistema (OREALC, 2013: 201-203) distingue los siguientes tipos de discapacidad:

- a) Discapacidad motora: limitaciones permanentes del sistema neuromuscular (posturales, de desplazamiento, coordinación de movimientos, expresión oral), debidas a un deficiente funcionamiento en el sistema óseo-articular, muscular o nervioso, y que limitan la capacidad funcional en grados muy variables. Las más habituales son la parálisis cerebral, la espina bífida y las distrofias musculares. La eliminación de barreras físicas y la provisión de ayudas técnicas, equipamientos y sistemas de comunicación alternativos o aumentativos son cruciales para facilitar la autonomía, la movilidad, la comunicación y el aprendizaje de los estudiantes con discapacidad motora.

- b) Discapacidad intelectual: se origina antes de los 18 años y se caracteriza por limitaciones significativas tanto en el funcionamiento intelectual como en la conducta adaptativa.
- Funcionamiento intelectual: se refiere a la capacidad mental general, como el aprendizaje, razonamiento y resolución de problemas, entre otros.
  - Conducta adaptativa: se compone de tres tipos de habilidades, conceptuales, sociales e interpersonales y prácticas.
- c) Discapacidades sensoriales, que distingue tres tipos:
- Discapacidad auditiva: se refiere a la disminución de la función auditiva en diferentes grados que tiene implicaciones en el desarrollo comunicativo, social y el aprendizaje de la lengua escrita. Desde una perspectiva sociocultural las personas sordas constituyen una comunidad que comparte una lengua y un código de conductas y valores que se aprenden y transmiten de una generación a otra. En el caso de las personas sordas es necesario el desarrollo de una educación bilingüe; lenguaje de señas como primera lengua y el aprendizaje de la lengua oral y escrita.
  - Discapacidad visual: limitación de la función visual que se caracteriza por una amplia gama de grados de visión, debida a causas congénitas o adquiridas. Suelen establecerse dos grandes categorías: la ceguera, pérdida total de la visión o ligera percepción la luz, y la baja visión, en la que existe un resto visual suficiente para ver la luz, orientarse por ella y emplearla con propósitos funcionales.
  - Las dificultades visuales tienen implicaciones en el acceso a la información, la orientación y los desplazamientos, por lo que las ayudas y los apoyos han de orientarse al aprendizaje del sistema braille, la estimulación visual y la movilidad, proporcionando a las personas con discapacidad visual los equipamientos y las herramientas informáticas que faciliten estos procesos.
  - Sordo-ceguera: es una discapacidad sensorial dual, que involucra una disminución significativa de la visión y de la audición, afectando exponencialmente otros aspectos del desarrollo, de manera distinta a la alteración que podría provocar la pérdida de visión o de audición por separado.
- d) Multidiscapacidad o retos múltiples: esto ocurre cuando se presenta más de un tipo de discapacidad en una persona. Los estudiantes con multidiscapacidad requieren para su adecuada atención apoyos extensos y generalizados en las áreas de comunicación, orientación, movilidad, vida diaria, socialización y en el aprendizaje en general.
- e) Trastornos generalizados del desarrollo: se caracterizan por alteraciones cualitativas en la interacción social recíproca, la comunicación verbal y no verbal y por un repertorio repetitivo, estereotipado y restrictivo de intereses y actividades.
- f) Otro tipo de discapacidad: esta categoría se incluye para que los países puedan incorporar información disponible no contenida en las categorías definidas o no plenamente coincidentes. La especificación solicitada para toda información que se reporte en ella permitirá su exhaustivo análisis así como considerar dicha información en función de si es acorde con la población objeto del sistema.

### 6.3. PROCESO DE DISEÑO DEL SIRIED

El diseño del Sistema se caracteriza por la interacción intensa y continua entre el equipo de la OREALC/UNESCO Santiago responsable de su formulación y los responsables en los países latinoamericanos de la educación especial, así como por la participación fundamental de especialistas en discapacidad, necesidades educativas especiales y educación inclusiva. El proyecto se desarrolló en tres fases (Blanco *et al.*, 2010):

- La primera consistió en la definición preliminar del Sistema de Información, estableciendo el marco conceptual-analítico, la cobertura y alcance, y la propuesta de un conjunto de variables a relevar y de indicadores a construir.
- En la segunda la propuesta inicial del SIRIED, se sometió a juicio de expertos y a la validación de los indicadores en cinco países de la región, elaborando a partir de sus resultados una versión más refinada, incluidas las herramientas de instrumentación.
- En la tercera fase, se realizó una validación local (en instituciones educativas) de tres países participantes en la fase previa, a partir de la cual se elaboró la versión definitiva del Sistema.

#### Fase 1

Para el cumplimiento del primer componente de la primera fase se realizó una encuesta con el objetivo principal de indagar en los países de Latinoamérica la disponibilidad y características del acopio de información estadística relacionada con las necesidades educativas especiales. Como objetivo secundario se indagó sobre la información que los países consideran apropiado y relevante disponer para monitorear la situación educativa de la población con necesidades educativas especiales y la formulación y seguimiento de políticas activas para ellos (SIRNEE, 2007).

Se solicitó a los 19 países de la región la designación de dos representantes: uno del área de Educación Especial y otro del Sistema de Información. De los 13 países que respondieron la encuesta, 10 adjuntaron además material adicional con definiciones, legislación y otros documentos relacionados (SIRNEE, 2007).

Los resultados pusieron de manifiesto la heterogeneidad de definiciones y clasificaciones existentes en la región y la complejidad y variedad de servicios orientados a la atención de los alumnos con necesidades educativas especiales. La encuesta también mostró las asimetrías en la cantidad, calidad e integridad de la información estadística disponible, debidas a las diferencias en el desarrollo y calidad de los sistemas de información educativa de los países de la región. En varios de ellos, los alumnos con necesidades educativas especiales se definen como aquellos que tienen mayores dificultades para aprender y requieren adaptaciones en el currículo o medios, recursos y apoyos especializados. Esta amplitud del concepto y la necesidad de acudir a otro, como el de dificultades de aprendizaje, hacen problemática la operacionalización de las necesidades educativas especiales. Otro aspecto debatido fue la vinculación entre la población objetivo del sistema de información y la población objetivo de la educación especial y de otras áreas de los ministerios de educación.

A partir de la puesta en común de los resultados con los directores de educación básica y educación especial de los países de la región, y tras debatir y acordar el foco y alcance del Sistema,

el equipo de la OREALC/UNESCO Santiago desarrolló el marco que fundamenta el modelo de análisis que se diseñó, junto con sus dimensiones, categorías e indicadores.

## Fase 2

La segunda fase tuvo como propósito validar el diseño preliminar del SIRIED con objeto de perfeccionarlo. Después de pasar por un comité de especialistas en indicadores educativos y educación especial, se piloteó en cinco países: Argentina, Brasil, Costa Rica, Guatemala y República Dominicana.

Este ejercicio permitió identificar la disponibilidad de la información, afinar la población objetivo, realizar adecuaciones a los marcos conceptual y metodológico, entre otras acciones de perfeccionamiento.

La reformulación del Sistema se puso a consideración de todos los representantes de la educación especial de los países de la región, lo que derivó en la creación de nuevos indicadores que proveerían información necesaria para la toma de decisiones (Blanco *et al.*, 2010).

## Fase 3

El objetivo fundamental de esta tercera fase fue validar el sistema de información a nivel local. Se consideró clave vincular su viabilidad en el nivel macro con su factibilidad en el nivel micro de las instituciones educativas. El objetivo de esta fase fue generar información que contribuyera a validar los distintos componentes del sistema y refinar la formulación de los indicadores y el diseño de las herramientas de acopio de la información.

Se seleccionaron dos países con realidades muy distintas, Argentina y Guatemala, y posteriormente se sumó Brasil, que asumió el financiamiento de la aplicación piloto, por considerar de gran relevancia este sistema de información. La validación local fue realizada por instituciones especializadas de cada país. Se aplicaron unos cuestionarios para obtener información cualitativa y cuantitativa en una muestra de establecimientos de educación regular, especial y de jóvenes y adultos de cada país, que representaban diferentes realidades en cuanto a la ubicación geográfica, tipo de gestión y niveles educativos. En las entrevistas se indagaron entre otros aspectos: la disponibilidad de la información estadística que el sistema necesita, y características de su registro y confiabilidad; el grado de apropiación de la normativa vigente y su concreción en el proyecto educativo institucional y prácticas pedagógicas; así como la visión de los docentes en relación con la educación inclusiva, indagando sobre las barreras para su aplicabilidad.

Al final de estas tres fases de diseño del proyecto, se configuró el modelo que opera el SIRIED en la actualidad, el cual se presenta en el siguiente subapartado (Blanco *et al.*, 2010).

## 6.4. MODELO DE ANÁLISIS PARA LA OPERACIÓN DEL SIRIED

La parte nodal del Sistema Regional de Información Educativa de los Estudiantes con Discapacidad (SIRIED) lo conforman las dimensiones, categorías e indicadores sobre los que se busca

sistematizar la información, es decir, su modelo analítico de operación. Estos elementos están en consonancia con el marco conceptual de la calidad de la educación, educación inclusiva y de atención a la diversidad, construido a propósito de los alumnos con discapacidad. Así, el modelo operativo del SIRIED está organizado bajo cinco dimensiones: relevancia, pertinencia, equidad, eficacia y eficiencia. En cada una de las dimensiones de la calidad, se desglosan categorías de análisis (véase tabla 15), las cuales a su vez contemplan un total de 42 indicadores cuantitativos y cualitativos (Blanco *et al.*, 2010).

Los indicadores cualitativos se orientan a conocer en qué medida la legislación educativa, los marcos curriculares y las medidas adoptadas por los países promueven y facilitan el desarrollo de sistemas educativos inclusivos, que aseguren el derecho de todas las personas, incluidas aquellas con discapacidad, a una educación de calidad en igualdad de condiciones. El análisis cualitativo de estos indicadores se complementa con un abordaje cuantitativo, dado que para hacer un seguimiento a través del tiempo es aconsejable convertir las variables de tipo cualitativo a una escala cuantitativa. Cada descriptor se califica de acuerdo a una escala ordinal que varía entre 0 y 4; si el indicador obtiene el valor de cero, señala la inexistencia de marcos legislativos, normativos, curriculares, políticas o programas/acciones, según corresponda, vinculados a la temática que indaga el indicador. En cambio, si el indicador obtiene el valor máximo (4), indica la presencia explícita del aspecto a que hace referencia en los marcos legislativos, normativos, curriculares, de política o programas/acciones.

Los indicadores cuantitativos permiten, a través de medidas estadísticas simples como tasas, proporciones o porcentajes, tener una aproximación a la capacidad de respuesta de las políticas públicas y los sistemas educativos a las necesidades educativas de los estudiantes con discapacidad. Esto permite evidenciar las brechas que aún persisten en el cumplimiento pleno del derecho a una educación inclusiva y de calidad, y cuantificar los esfuerzos que es preciso realizar para su pleno cumplimiento (Blanco *et al.*, 2010).

Cada uno de los indicadores del SIRIED se define de forma operacional en una matriz (véase tabla 16), de tal forma que los países dispongan de una descripción precisa y compartida sobre lo que se pretende medir con cada uno de ellos, para que con posterioridad se definan las fuentes de donde se obtendrá la información necesaria, o bien, se prevea la creación de los mecanismos para generarla.

El diseño del SIRIED, al tener un alcance regional, se ha basado en las pautas técnico-metodológicas del Instituto de Estadística de la UNESCO (UIS), la Clasificación Internacional Normalizada de la Educación en su versión de 1997, la Clasificación Internacional de Enfermedades de la Organización Mundial de la Salud y la Clasificación Internacional del Funcionamiento, de la Discapacidad y la Salud, que en conjunto, permiten la integración de estadísticas provenientes de otros subsistemas sociales y garantizan la comparabilidad regional.

Es importante advertir que el Sistema se concibe a sí mismo en constante evolución, pues admite la incorporación de nuevos indicadores y, en determinado momento, la eliminación de otros, como resultado de la escasa utilidad que lleguen a tener para la toma de decisiones, ante la superación del fenómeno sobre el que se pretendía informar mediante esos indicadores (OREALC, 2012).

**Tabla 15** Dimensiones y categorías del SIRIED

Dimensión	Categoría	Descripción
Relevancia	Fines y contenidos de la educación	Esta categoría se propone indagar en qué medida los fines de la educación y el currículo común son el referente para la educación de los estudiantes con discapacidad, independientemente de dónde estén escolarizados. Este aspecto es especialmente relevante en el caso de los estudiantes con discapacidad porque en algunos países todavía existen currículos paralelos para ciertos tipos de discapacidad o para los centros de educación especial.
	Atención a la diversidad	Se refiere a los aspectos normativos vinculados con el respeto y valoración de la diversidad, y de forma particular con el derecho de los estudiantes con discapacidad a una educación de calidad y a ser educados junto a sus pares en las escuelas regulares de su comunidad.
Pertinencia	Accesibilidad / adaptabilidad	La accesibilidad y la adaptabilidad son dos de los parámetros para evaluar el cumplimiento del derecho a la educación, y son indispensables para asegurar el acceso, la participación y el aprendizaje de los estudiantes con discapacidad en igualdad de condiciones con los demás. El principio de accesibilidad implica el diseño de productos, entornos, programas y servicios que puedan utilizar todas las personas, en la mayor medida posible, sin necesidad de adaptación ni diseño especializado. La accesibilidad no se refiere solamente al acceso físico, sino también a la información, la comunicación, el currículo y las actividades educativas.
	Servicios de apoyo	Definidos como todos aquellos recursos humanos que complementan o refuerzan la acción pedagógica de los docentes para dar respuesta a las necesidades educativas de todos los estudiantes, prestando especial atención a aquellos que más ayuda necesitan para optimizar su desarrollo, así como para participar y avanzar en su aprendizaje. Estos apoyos constituyen un continuo que puede ir desde el apoyo entre estudiantes, docentes, y familias de las propias escuelas, a recursos de apoyo especializado, como profesores, orientadores, psicólogos, fonoaudiólogos o equipos de carácter interdisciplinario.
Equidad	Igualdad de oportunidades	Desde la perspectiva de la equidad, es preciso equilibrar los principios de igualdad (lo común) y de diferenciación (lo diverso). La educación debe tratar de forma diferenciada lo que es desigual en el origen para llegar a resultados de aprendizaje equiparables y no reproducir las desigualdades presentes en la sociedad. Un trato diferenciado que no sea discriminatorio o excluyente significa no sólo dar más a quién menos tiene, sino también dar a cada quien los recursos que requiere, es decir, personalizar las ayudas.
	Disparidad según características sociodemográficas	Esta categoría se refiere a la identificación de disparidades al interior de la población con discapacidad que denotan elementos discriminatorios en lo que atañe al ejercicio del derecho a la educación. Se consideran disparidades por género, zona de residencia y pertenencia étnica.
Eficacia	Acceso y conclusión de estudios	Una de las condiciones básicas para hacer efectivo el derecho a la educación es el acceso a los servicios educativos como inicio de la trayectoria educativa. La posibilidad de contar con información de los niveles de acceso al sistema educativo permite dar cuenta de qué proporción de la población está siendo atendida y, como complemento, el volumen de población que se encuentra excluida. El acceso es una condición necesaria pero insuficiente para asegurar el derecho a la educación, ya que no garantiza por sí mismo la permanencia en el sistema educativo, el aprendizaje y la culminación de la educación obligatoria. La conclusión de estudios se refiere al egreso y la certificación de un determinado nivel educativo (primaria, secundaria o terciaria).
	Docentes	La calidad de los docentes y el ambiente que generan en la sala de clase, excluidas las variables extraescolares, son los factores más importantes que explican los resultados de aprendizaje de los estudiantes. Los países de la región enfrentan el desafío de implementar políticas y estrategias que garanticen que los docentes cuenten con las competencias profesionales y éticas adecuadas, y los medios necesarios para hacer efectivo el derecho de todos los estudiantes a aprender y desarrollarse plenamente. A su vez, la educación inclusiva requiere la formación y contratación de docentes representativos de la diversidad presente en la sociedad.
Eficiencia	Trayectorias educativas	Uno de los objetivos del sistema de información es conocer la situación educativa de los estudiantes con discapacidad, lo cual significa obtener información no sólo sobre su acceso a la educación, sino también sobre su biografía escolar (repetición, atraso escolar, ingreso tardío, abandono escolar).
	Gestión institucional	Avanzar hacia una educación de calidad para todos exige un nuevo modelo de escuela y el desarrollo de políticas que faciliten su puesta en práctica. Se requieren cambios en la cultura, organización y prácticas educativas de las escuelas que permitan tanto la participación como el aprendizaje de todos los estudiantes, lo mismo que la construcción de comunidades de aprendizaje y de colaboración, así como la apertura de las escuelas a las familias y la comunidad.

Fuente: Blanco *et al.*, 2010, pp. 36-42.



**Tabla 16** Matriz de indicadores del SIRIED

Dimensión	Categoría	Indicadores
Relevancia	Fines y contenidos de la educación	1. Se reconoce que los fines y objetivos generales de la educación son los mismos para todos los estudiantes.
		2. El currículo común es el referente para la educación de todos los estudiantes sin ningún tipo de discriminación.
		3. Existe un currículo con una orientación inclusiva que considera la diversidad de necesidades de aprendizaje de todos los estudiantes.
Atención a la diversidad		4. Se garantiza el derecho a la educación de las personas con discapacidad en igualdad de condiciones con los demás.
		5. Se asegura el derecho de las personas con discapacidad a acceder a una educación inclusiva en todos los niveles de enseñanza.
		6. Se promueve el acceso de los niños y las niñas con discapacidad a los servicios y programas de atención y educación de la primera infancia.
		7. Se promueve el desarrollo de propuestas metodológicas que faciliten el aprendizaje y la participación de todos los estudiantes.
		8. En los procedimientos de evaluación y promoción se consideran las necesidades y características particulares de los estudiantes con discapacidad.
		9. Porcentaje de estudiantes con discapacidad matriculados en la educación regular, por nivel educativo.
		10. Número de organizaciones de la sociedad civil que participan en la promoción y protección del derecho a la educación de las personas con discapacidad.
Pertinencia		11. Se garantiza el aprendizaje de códigos de comunicación aumentativos, alternativos o complementarios al lenguaje oral y escrito para asegurar el aprendizaje y la participación de los estudiantes con discapacidad.
		12. Existen criterios y orientaciones para el diseño universal en la construcción de edificios escolares.
		13. Porcentaje de edificios escolares con diseño universal o adaptaciones en su infraestructura que garanticen la accesibilidad física a todos.
		14. Porcentaje de estudiantes con discapacidad que reciben los equipamientos y materiales específicos que necesitan para su participación y accesibilidad al currículo, por nivel educativo.
		15. Porcentaje de estudiantes con discapacidad que acceden a los códigos de comunicación complementarios o alternativos al lenguaje oral y escrito que necesitan para garantizar el acceso a la información, la participación y el aprendizaje, por nivel educativo.
		16. Se definen procedimientos para identificar las barreras y las necesidades de apoyo y de recursos que requieren los estudiantes con discapacidad.
Servicios de apoyo		17. Existen servicios de apoyo con una orientación inclusiva para favorecer el aprendizaje y la participación de todos los estudiantes.
		18. Porcentaje de establecimientos de educación regular que incluyen estudiantes con discapacidad y reciben servicios de apoyo por nivel educativo.

Dimensión	Categoría	Indicadores
Equidad	Igualdad de oportunidades	19. Existen políticas intersectoriales para garantizar la igualdad de condiciones de las personas con discapacidad en el pleno ejercicio del derecho a la educación.
		20. Se identifican las barreras económicas y se adoptan políticas y medidas para garantizar la gratuidad de la educación para las personas con discapacidad.
		21. Porcentaje de estudiantes con discapacidad beneficiarios de servicio de transporte, becas y alimentación, según tipo de discapacidad.
		22. Porcentaje de establecimientos de educación regular que atienden estudiantes con discapacidad, por tipo de gestión y nivel educativo.
		23. Porcentaje de establecimientos de educación de jóvenes y adultos que atienden estudiantes con discapacidad, por tipo de gestión y nivel educativo.
Equidad	Disparidad según características sociodemográficas	24. Tasa específica de matrícula por edad de la población con discapacidad de 0 a 24 años.
		25. Distribución porcentual según género de los y las estudiantes con discapacidad, por tipo de discapacidad.
		26. Relación entre los estudiantes con discapacidad según género y el total de varones y mujeres matriculados en el sistema, por nivel educativo.
		27. Relación entre los estudiantes con discapacidad y el total de estudiantes matriculados en las zonas urbana y rural, por nivel educativo.
Eficacia	Acceso y conclusión de estudios	28. Relación entre los estudiantes matriculados con discapacidad por cada mil estudiantes matriculados en el sistema educativo según pertenencia étnica y género, por nivel educativo.
		29. Número de estudiantes con discapacidad, por cada mil estudiantes matriculados en el sistema educativo, por nivel educativo y tipo de discapacidad.
		30. Número de estudiantes con discapacidad por cada mil estudiantes matriculados en el sistema educativo formal y no convencional, por edad simple para el grupo de 0 a 6 años.
	Docentes	31. Número de graduados con discapacidad por cada mil graduados del sistema educativo, por nivel educativo y tipo de discapacidad.
		32. Porcentaje de estudiantes con discapacidad aprobados, reprobados y que abandonaron el sistema educativo, por nivel educativo y tipo de discapacidad.
		33. Porcentaje de docentes que reciben formación continua relacionada con la educación inclusiva, la atención a la diversidad y las necesidades educativas de los estudiantes con discapacidad, por nivel educativo.
Eficiencia	Trayectorias educativas	34. Porcentaje de docentes con discapacidad, por nivel educativo y tipo de discapacidad.
		35. La formación de los docentes incluye competencias para la atención a la diversidad y el desarrollo de la educación inclusiva.
		36. Porcentaje de estudiantes con discapacidad matriculados en educación regular, en condición de repetidores, por nivel educativo.
	Gestión institucional	37. Estudiantes con discapacidad matriculados en educación regular, atrasados 2 o más grados.
		38. Porcentaje de estudiantes con discapacidad que se matriculan en centros de educación especial después de haber transitado por la educación regular, según el tipo de discapacidad.
		39. Se promueve el desarrollo de proyectos educativos institucionales con una orientación inclusiva.
		40. Se garantiza la participación de todos los estudiantes en la toma de decisiones.
		41. Se establecen instancias para la participación de las familias.
		42. Se promueve la participación de la comunidad para atender la diversidad de los estudiantes.

Fuente: Blanco *et al.*, 2010, pp. 45-47.

## 6.5. PRIMEROS RESULTADOS DE LA IMPLEMENTACIÓN DEL SISTEMA

En 2013 se publicaron los primeros resultados de la aplicación del SIRIED realizada en 2010, en 8 de los 19 países de la región: Argentina, Brasil, Costa Rica, Guatemala, México, Paraguay, Perú y República Dominicana (OREALC, 2013). Su implementación es gradual tanto en la cantidad de países como en los indicadores sobre los que se ha ido incorporando información. Así, en la primera fase de aplicación se solicitó información sobre 14 de los 42 indicadores que incluye todo el Sistema.

Los indicadores sobre los que se obtuvo información en la primera aplicación fueron:

- Porcentaje de estudiantes con discapacidad matriculados en educación regular, por nivel educativo (indicador 9)
- Porcentaje de establecimientos de educación regular que atienden estudiantes con discapacidad, por tipo de gestión y nivel educativo (indicador 22)
- Porcentaje de estudiantes mujeres/varones con discapacidad matriculados con relación a la matrícula total de mujeres/varones (indicador 26)
- Relación entre los estudiantes con discapacidad y el total de estudiantes matriculados en las zonas rural y urbana, por nivel educativo (indicador 27)
- Número de estudiantes con discapacidad por cada mil estudiantes en el sistema educativo, por nivel educativo, zona, tipo de gestión y sexo (indicador 29)

Los indicadores cuantitativos se desglosan por el tipo de sostenimiento de la escuela y su ubicación (rural/urbana), lo que permite diseñar políticas públicas más enfocadas en ciertos subgrupos de personas con discapacidad.

Los resultados pueden resumirse de la siguiente manera:

- Los ocho países disponen de legislación o normatividad que prohíbe y sanciona la discriminación; en la mayoría incluso, este elemento forma parte de su constitución nacional. La mención específica a la discriminación de las personas por motivos de discapacidad se encuentra por lo general en las leyes de discapacidad.
- El esfuerzo que realizan los países por incorporar al sistema educativo a la población con discapacidad es muy importante. En 2010, en los siete países para los que se presentan estadísticas, el sistema educativo formal incluyó a 1 140 478 estudiantes con discapacidad. De ese total, 69.4% está matriculado en escuelas regulares, y 30.6% en escuelas especiales.
- En la mayoría de los países puede afirmarse que los estudiantes con discapacidad registran un mayor acceso al nivel primario (sin poder cuantificar cuántos no acceden al mismo), un acceso más restringido a preescolar y un bajo nivel de acceso a secundaria, llegando una ínfima porción a cursar el equivalente de la educación media superior en México.
- Todos los países han avanzado en la implementación de servicios de apoyo indispensables para asegurar la igualdad de oportunidades de los estudiantes con dificultades y garantizar el máximo desarrollo, sin que sus deficiencias se constituyan en impedimentos para ello. Se reconoce la necesidad de avanzar en la construcción de indicadores cuantitativos que permitan constatar cuántos estudiantes requieren cada uno de los distintos tipos de apoyo y cuántos son verdaderos receptores de ellos (OREALC, 2013).

El SIRIED es un esfuerzo encomiable en la región que evidencia la preocupación por hacer más visibles a subgrupos vulnerables, tales como las personas con alguna discapacidad, para hacer efectivo el derecho a una educación de calidad.

Más allá de lo valioso de los primeros resultados, en el INEE se podrían recuperar los marcos de actuación en los que se apoya el SIRIED así como la metodología para la construcción del sistema y sus indicadores. La interacción frecuente desde la concepción misma del proyecto entre sus diseñadores, tomadores de decisiones, estudiosos del fenómeno y actores educativos que trabajan directamente con las subpoblaciones de interés es una estrategia que fortaleció al Sistema.

# 7

## Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)

### 7.1. DESCRIPCIÓN GENERAL

En el continente africano los sistemas nacionales de evaluación educativa son escasos, y los que existen suelen ubicarse en un nivel de madurez más bien en ciernes. En los sistemas nacionales de monitoreo y evaluación educativa el foco de interés todavía está puesto en el acceso a la educación obligatoria (cobertura), las tasas de reprobación (que son muy altas), la deserción, la inasistencia de los maestros y alumnos, la disponibilidad de espacios físicos educativos seguros e higiénicos así como de materiales de enseñanza y, sólo en algunos casos se levanta información sobre el logro de los estudiantes en Lengua y Matemáticas que en no pocas ocasiones se ha derivado de las evaluaciones internacionales como el PISA o el Estudio de Tendencias Internacionales en Matemáticas y Ciencia (TIMSS, por sus siglas en inglés).

Al disponer de sistemas de monitoreo educativo aún en fases iniciales de desarrollo, no hay mucha evidencia del trabajo técnico que se realiza en ellos, y entre los materiales que se identificaron está prácticamente ausente la consideración de la diversidad en los procesos evaluativos. Si bien en las naciones africanas se atiende la diversidad en la operación del sistema educativo, no ocurre lo mismo en los sistemas de monitoreo o evaluación educativa.

Pese a esta carencia a nivel nacional, el Consorcio para Monitorear la Calidad Educativa en el Sur y Este de África (SACMEQ, por sus siglas en inglés), organismo regional, sí cuenta con algunos desarrollos en materia de evaluación inclusiva, lo que destaca en el contexto del continente africano; en particular, es clara la preocupación por la equidad de género en el ámbito educativo.

El SACMEQ, surgido en 1995 con el apoyo técnico del Instituto Internacional de Planeación Educativa (IIEP, por sus siglas en inglés) de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, por sus siglas en inglés) y con apoyo financiero del gobierno de Holanda, es una organización internacional sin fines de lucro para el desarrollo de 15 Ministerios de Educación de África meridional y oriental (correspondientes a 14 países), que decidieron trabajar juntos para compartir experiencias y conocimientos en el desarrollo de las capacidades de los planificadores de la educación para aplicar métodos científicos de monitoreo y evaluación de las condiciones de escolarización y la calidad de la educación (Murimba, 2005).

El principal objetivo del trabajo del SACMEQ ha sido implementar programas de investigación y formación para dotar a los planificadores de la educación e investigadores con conocimientos técnicos pertinentes sobre diseño de investigación, construcción de instrumentos, muestreo, análisis de datos y redacción de informes.

Los 14 países que conforman el Consorcio son: Angola, Botswana, Kenia, Lesoto, Malawi, Mauricio, Mozambique, Namibia, Seychelles, Sudáfrica, Suazilandia, Tanzania (que tiene dos ministerios de educación: Mainland y Zanzibar), Uganda, Zambia y Zimbabwe.

El SACMEQ desarrolla estudios transnacionales de investigación en gran escala en los países miembros para evaluar las condiciones de escolarización y los niveles de rendimiento de los alumnos y maestros en áreas de Lengua, Matemáticas, así como niveles de conocimiento por parte de los educandos sobre salud elemental. El SACMEQ ha desarrollado y concluido tres proyectos de investigación (llamados SACMEQ I, II y III) y en la actualidad se está implementando el cuarto proyecto, del que aún no se han reportado resultados ni los procedimientos empleados.

El Consorcio realiza sus proyectos de forma holística con los tomadores de decisiones en materia educativa de los 14 países miembros. Así, los propósitos y las variables que se derivan de dichos proyectos, se han ido adaptando en función de las necesidades de información de los diseñadores de las políticas públicas educativas. Por eso mismo, cada proyecto del SACMEQ se ha vuelto más ambicioso e incorpora cada vez una mayor cantidad de aspectos a indagar. Ante la falta de información detallada sobre el último de los proyectos del Consorcio, en este trabajo se privilegia la presentación del tercero.

Los aspectos indagados más importantes del SACMEQ III fueron: *a)* las condiciones generales de escolarización, *b)* los niveles de logro en lectura y matemáticas de los alumnos de 6° grado y sus maestros, y *c)* los conocimientos que los alumnos y sus maestros tienen sobre el Virus de Inmunodeficiencia Humana (VIH) y el Síndrome de Inmunodeficiencia Adquirida (SIDA).

En el proyecto SACMEQ III se emplearon muestras representativas a nivel nacional, construidas mediante un muestreo bietápico y estratificado. En la primera etapa se seleccionaron escuelas de acuerdo a su tamaño (PPT, probabilidad proporcional al tamaño), utilizando como estrato la provincia a la que pertenecen; en la segunda etapa de muestreo se seleccionó aleatoriamente a 25 alumnos dentro de cada escuela escogida en la etapa anterior (Wasanga, Ogle y Wambua, 2012).

La aplicación se realizó en 2007, en total se obtuvo información de 61 396 alumnos ubicados en 2 779 escuelas, que eran atendidos por 8 026 profesores (Hungu, 2011b), distribuidos todos ellos en los 14 países participantes.

La población objetivo del proyecto III fueron los alumnos de sexto grado de escuelas regulares. De la población a muestrear se excluyeron las escuelas de educación especial y los centros escolares con menos de 15-20 alumnos de sexto grado, lo que sin duda es una práctica no deseable de acuerdo al marco conceptual de este informe, esto es, la de extraer de los operativos de evaluación subpoblaciones vulnerables, que en este caso son aquellas que se ubican en localidades aisladas o de contexto rural, o enfrentan alguna condición de desventaja respecto al resto de la población.

Para recolectar la información se aplicaron tres pruebas: Lengua, Matemáticas y conocimiento sobre VIH y SIDA, así como cuatro cuestionarios de contexto: para alumnos, padres, profesores y directores de los planteles muestreados. Los ítems de las pruebas de Lengua y Matemáticas fueron prácticamente los mismos que los utilizados en la segunda versión del proyecto; en cambio, los de la tercera prueba se tuvieron que generar por tratarse de un dominio de reciente creación. Estos reactivos se pilotearon en Botswana previo a su aplicación definitiva;

el pilotaje se hizo para asegurar que el lenguaje utilizado en los reactivos fuera accesible para los alumnos de 6° y evitar así algún sesgo cultural y, además, para pilotear si los alumnos podían seguir las instrucciones indicadas.

Al tratarse de una evaluación en gran escala, se estandarizó la aplicación de los instrumentos con el apoyo de una capacitación muy cuidadosa para aplicadores (SACMEQ, 1995a, 1995b), al igual que el proceso de manejo de la información en bases de datos, el que se inició con un proceso de limpieza y depuración (SACMEQ, 2008).

En lo que corresponde al análisis, si bien se realizaron algunos particulares en cada país (Jopo, Maema y Ramokoena, 2011b; Leste y Benstrong, 2010; Mhonyiwa, Ponera y Mrutu, 2011b; Milner G., Mulera D. y Chimuzu T., 2011; Miranda, Amadhila, Dengeinge y Shikongo, 2011; Mohammed, Mohamed, Abdalla, Masoud y Sharif, 2010; Meshack Qetelo Moloji y Chetty, 2010; Monyaku y Mmerekí, 2010; Musonda y Kaba, 2010; Sauba y Lutchimiah, 2010; Wasanga *et al.*, 2012), los análisis básicos incluyeron los niveles de logro en cada una de las tres pruebas y su relación con las variables de contexto indagadas mediante los cuestionarios aplicados.

A nivel de estudiante, los factores que más intervinieron en el logro académico fueron la repetición de grado, el contexto socioeconómico, la edad y el sexo; mientras que, a nivel de escuela, las que tuvieron un mayor efecto en el rendimiento de los alumnos fueron los recursos escolares y la ubicación de la escuela (Hungí, 2011a, 2011a).

## 7.2. PRÁCTICAS DE ATENCIÓN A LA DIVERSIDAD

En el proyecto SACMEQ III se denota una preocupación constante por la atención a la equidad de género, de manera especial en las fases de análisis, interpretación y difusión de los resultados. En el contexto africano esta preocupación tiene mucho sentido por la discriminación de que han sido objeto las mujeres, tanto en el desarrollo social en general como en el educativo en particular.

La equidad de género en educación es definida en la misma forma en que lo hace la UNESCO, esto es, que hombres y mujeres experimenten las mismas ventajas y desventajas al asistir a la escuela, y reciban los mismos métodos de enseñanza, currículo y orientación académica, para producir iguales logros educativos y subsecuentes oportunidades para la vida (Amadhila, Dengeinge, Miranda y Shikongo, 2011).

Por ello los reportes de evaluación incluyen de manera sistemática el desglose de resultados de acuerdo al género, de las variables analizadas, ya sea en términos porcentuales o promedios aritméticos, lo que permite identificar el grado de disparidad entre sexos para cada uno de los rubros evaluados, rendimiento académico en Lengua y Matemáticas, conocimiento sobre VIH/SIDA y condiciones generales de escolarización.

La preocupación por la equidad de género no sólo se advierte en las evaluaciones de los alumnos. Un estudio secundario de datos llevado a cabo a partir de la información del SACMEQ III da cuenta de los desbalances que prevalecen en la gran mayoría de los 14 países respecto al sexo de docentes y directivos escolares, siendo el masculino el predominante (Hungí, 2010, 2011c; Makuwa, 2011).

Así, la consideración de la diversidad de género es una línea de análisis que subyace en los estudios que se han generado a partir de los datos de logro y de contexto. Una evidencia de ello es la elaboración de reportes regionales (Dolata, 2011; Saito, 2004, 2010a, 2010b, 2011) y breviaros de política educativa por cada nación, dedicados específicamente a mostrar cómo la variable sexo (controlando otras variables contextuales) interviene en los niveles de logro en Lengua, Matemáticas y conocimiento sobre el VIH y SIDA. Los breviaros de política de cada país, corresponden a Botsuana (Monyaku y Mmerekí, 2011), Kenia (Wasanga, Ogle y Wambua, 2011), Lesoto (Jopo, Maema y Ramokoena, 2011a), Malawi (Milner, Mulera, Banda, Matale y Chimombo, 2011), Mauricio (Sauba y Lutchimiah, 2011), Mozambique (Passos, Nahara y Magaia, 2012), Namibia (Amadhila *et al.*, 2011), Seychelles (Leste y Benstrong, 2011), Suazilandia (Shabalala y Nxumalo, 2011), Sudáfrica (Meshack Q. Moloi y Chetty, 2011), Tanzania (Mhonyiwa, Ponera y Mrutu, 2011a), Uganda (Saito, Ssenbulya y Lubega, 2011), Zambia (Kaba y Musonda, 2011) y Zimbabue (Chitiga y Chinoona, 2011).

Los breviaros de política están destinados a tomadores de decisiones de cada uno de los países que forman parte del Consorcio. En cada uno de los textos se presenta información sobre el acceso diferenciado de hombres y mujeres a la educación básica, el desbalance en la matriculación de las mujeres con respecto a los hombres y el avance que ha mostrado este indicador entre 2000 y 2007, años en que se aplicaron los proyectos SACMEQ II y III, respectivamente; asimismo, se incluyen las diferencias por género en el logro obtenido en las pruebas de Lengua y Matemáticas. Además, en los breviaros se incluye una descripción sobre el personal femenino que labora en las escuelas, la seguridad escolar y aspectos relacionados con la higiene, todos ellos con especial atención a las mujeres. Al final de cada texto se ofrecen orientaciones y sugerencias de política pública, todas ellas encaminadas a disminuir las desigualdades educativas entre niños y niñas, así como entre el personal docente y directivo que los atiende.

Pese al impulso que el SACMEQ le ha dado a la consideración de la diversidad de género en las evaluaciones, aún falta un mayor desarrollo. En la etapa de diseño de los instrumentos, si bien se realiza un pilotaje de los reactivos para validar la diversidad lingüística, la diversidad de género no es un elemento que se tome en cuenta. Asimismo, los análisis psicométricos no incluyen un análisis sobre el funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés) que explore un posible sesgo hacia las niñas o los niños, o bien, algún otro método que permita identificar esto.

Así, pese a que en los estudios del SACMEQ es patente la preocupación por una evaluación inclusiva de las mujeres, que se advierte en la presentación de resultados, la ausencia de estrategias de validez cultural desde el diseño de los instrumentos constituye un riesgo para la validez de las interpretaciones y usos de los resultados. Esta aseveración se fundamenta en el marco conceptual de este trabajo, específicamente, que una evaluación es justa cuando permite que los sujetos que participan en ella no se vean limitados para responderla por cuestiones ajenas al constructo que se pretende medir. En otras palabras, en las pruebas de Lengua y Matemáticas que se aplican en el SACMEQ se requiere asegurar que los resultados que obtienen las niñas sean producto del dominio que ellas tienen sobre Lengua y Matemáticas, y descartar que los reactivos contengan términos que le sean más familiares a los hombres por el tipo de actividades que ellos suelen realizar fuera de la escuela.

Cabe resaltar que son escasos los manuales técnicos o documentación de los procesos que se siguen para el desarrollo de los estudios del SACMEQ, lo que limitó la presentación detallada de los procedimientos que emplean.



### 7.3. OPINIÓN DE LOS MEDIOS

Se encontraron pocas notas informativas en Internet que hicieran alusión a los resultados que publica el SACMEQ y en ninguna de ellas se hacen consideraciones sobre la atención a la diversidad o se aprovechan los resultados desglosados por sexo que publica el Consorcio.

Las escasas notas identificadas dejan ver que la información proporcionada por SACMEQ se utiliza de forma complementaria a otras para hacer juicios sobre el sistema educativo de algún país en particular, principalmente Sudáfrica. Las notas en los medios tratan principalmente sobre el nivel que ocupa determinado país en lo relativo a la calidad de la educación, y se utilizan los resultados de las pruebas de logro que aplica el Consorcio para justificar o criticar el estado en que se encuentra la calidad educativa en el país sobre el que versa la nota.

Además de los ordenamientos de países, se identificaron dos aspectos que aparecen en varias ocasiones: las pruebas *Matric* que tienen que presentar los estudiantes de Sudáfrica para graduarse de la educación básica y el señalamiento de que los docentes más jóvenes son quienes consiguen que sus alumnos obtengan los mejores resultados. Sobre el primer aspecto, se utilizan los resultados de logro del SACMEQ para criticar la necesidad de aplicar las *Matric* si ya se dispone de los resultados del Consorcio, posición que es cuestionada señalando que cada una de las pruebas tiene propósitos distintos y, además, que las *Matric* son instrumentos censales para dar resultados por alumno, mientras que las evaluaciones del SACMEQ son muestrales para otorgar información a la tomadores de decisiones.

En el otro aspecto, los resultados de los alumnos atendidos por profesores jóvenes, se emplean los resultados del Consorcio y se hacen comparaciones entre los resultados de matemáticas de los alumnos según la edad de sus profesores; posteriormente se afirma que aquellos niños que son atendidos por maestros de entre 19 y 29 años consiguen en promedio puntajes más altos en matemáticas que sus contrapartes atendidos por docentes de 30 años o más.

Los artículos se encuentran en los anexos electrónicos. En la carpeta \SACMEQ\Medios

### 7.4. ESTUDIOS SOBRE LAS EVALUACIONES SACMEQ

No se identificaron metaevaluaciones del SACMEQ.



## Conclusiones

La revisión realizada en el marco de este trabajo sobre sistemas de evaluación educativa con enfoque de atención a la diversidad da cuenta que esta preocupación es relativamente reciente. A nivel internacional, surgió después de la segunda mitad del siglo XX. Este interés ha estado relacionado con la promoción de los derechos humanos y en particular con el derecho a la educación de los individuos.

La búsqueda de información relativa a las prácticas de atención a la diversidad en los diferentes sistemas educativos y programas de evaluación fue compleja por diversas razones. En primer lugar por la dependencia que se tuvo de las fuentes electrónicas para la identificación de las prácticas en cada uno de los sistemas estudiados. En segundo lugar, debido a que la conceptualización como atención a la diversidad no se encuentra enunciada como tal dentro de la documentación analizada; esto implicó una revisión general del proceso de desarrollo de las pruebas para la identificación de las prácticas que en los hechos pueden ser consideradas de atención a la diversidad. Y, en tercer lugar, por la escasez de estudios destinados a investigar o evaluar las prácticas implementadas en estos sistemas.

La muestra de prácticas en materia de atención a la diversidad de los países y organismos internacionales que aquí se presenta no debería asumirse como un referente hacia lo que se aspiraría que transitara en forma mecánica el Sistema Nacional de Evaluación Educativa en México. Las prácticas descritas ilustran una gama de posibilidades, pero es importante resaltar la complejidad del proceso de toma de decisiones en relación con la atención de la diversidad en las evaluaciones estandarizadas, que no puede basarse sin más en experiencias y decisiones propias de otros lugares, con procesos y condiciones específicas, no necesariamente coincidentes con las del sistema educativo del país.

A continuación se describen algunas comprensiones derivadas de la revisión de las prácticas internacionales en materia de consideración a la diversidad que pueden ser de especial utilidad para la conformación de decisiones institucionales en la materia por parte del Instituto Nacional para el Evaluación de la Educación (INEE):

1. En el proceso de análisis y toma de decisiones sobre las prácticas de atención a la diversidad en las evaluaciones, una de las etapas cruciales y más desafiantes es la definición de grupos y subgrupos que sean lo más homogéneos posible con respecto a sus necesidades.

Esto es importante, debido a que se ha encontrado que la eficacia de las acomodaciones varía de acuerdo con las características de los estudiantes (Pennock-Roman y Rivera, 2011: 11), y que no existe una acomodación que sea adecuada para todos los alumnos que pertenezcan a un grupo.

Por ejemplo, en un grupo de estudiantes que compartan la característica de que su primera lengua sea distinta a aquella en que se aplica la prueba, seguramente habrá diferencias en factores como el dominio que tengan de la lengua en la que se aplica la prueba, el dominio de la lengua materna, o de la lengua en la cual reciben la enseñanza, entre otros; si se implementa una sola acomodación para este grupo, su eficacia será distinta entre los subgrupos identificables. Esto ha sido mostrado a través de diferentes estudios; por ejemplo, en un metaanálisis desarrollado por Pennock-Roman y Rivera (2011) sobre la eficacia de las acomodaciones para alumnos cuya primera lengua era distinta al inglés, se clasificó a los alumnos de acuerdo a su nivel de dominio de la lengua en que serían evaluados. Inicialmente dividieron a los alumnos en dos grandes grupos: a) dominio bajo, b) dominio intermedio. Este último grupo fue subdividido después en dos grupos más: intermedio-bajo e intermedio-alto. La efectividad de las acomodaciones en estos grupos mostró las siguientes variaciones: para los alumnos con bajo dominio de la lengua inglesa que estaban recibiendo enseñanza en español, la acomodación más efectiva fue la versión en español de la prueba; para los alumnos con nivel de dominio intermedio-alto, la acomodación más efectiva fue la prueba en inglés simplificado o plano (comparada con la versión en español, o la versión que incluye la prueba en las dos lenguas); para los alumnos con nivel de dominio intermedio-bajo, el inglés plano tuvo un tamaño de efecto que no era significativamente distinto a cero, pero fue incluso mayor que el asociado a la versión en español de la prueba; e igual que la versión en dos lenguas de la prueba; para quienes reciben enseñanza en inglés, la versión en inglés plano no fue efectiva, pero fue mejor que la versión en español.

Es importante señalar que incluso se ha encontrado entre los alumnos que no son clasificados como hablantes de una lengua distinta a la de aplicación de la prueba, que puede haber heterogeneidad. En estos casos se sugiere también que se conformen subgrupos: hablantes nativos de la lengua en que se aplica la prueba; estudiantes que hablan lenguas de grupos minoritarios, pero que han sido clasificados con dominio de la lengua en que se aplica la prueba; y estudiantes que hablan lenguas de grupos minoritarios y que han salido de programas enfocados a alumnos cuya primera lengua no es aquella en la que se aplica la prueba (Pennock-Roman y Rivera, 2011).

2. La clasificación de los sustentantes debería diferenciar dos grupos de estudiantes: quienes tienen una discapacidad, y quienes tienen necesidades culturales en general, o lingüísticas en particular.

Algunas políticas de atención a la diversidad tienden a considerar a estos grupos como uno solo, y agrupan a los alumnos cuya lengua materna no es la que se usa en las pruebas con los alumnos con discapacidades. El problema de esta agrupación es que puede implicar una postura deficitaria acerca de las minorías lingüísticas (Rivera y Collum, 2004).

Se considera que los alumnos cuya primera lengua es distinta a aquella en que se aplica la prueba deben ser concebidos como susceptibles de una acomodación, debido a que, como aún están desarrollando su dominio de la lengua, es muy probable que los exámenes para medir conocimientos y habilidades de otras asignaturas, reflejen inadecuadamente sus logros (Pennock-Roman y Rivera, 2011). Sin embargo, la definición de los subgrupos es particularmente desafiante porque para la clasificación de los alumnos

de acuerdo al dominio de la lengua materna y la lengua en que se aplica la prueba, se depende por lo general de la clasificación que realizan las escuelas y los docentes, y estas clasificaciones varían entre escuelas, zonas y otros niveles de agregación de los sistemas educativos.

En relación con los alumnos que tienen alguna discapacidad, facilitar el acceso a las pruebas ha sido una prioridad en algunos países (*cfr.* caso de los Estados Unidos de Norteamérica). Las discapacidades pueden clasificarse en diferentes tipos: discapacidades de aprendizaje, limitaciones visuales o auditivas, discapacidades múltiples, discapacidades en lenguaje y comunicación, discapacidad emocional y de comportamiento, problemas de atención, y limitaciones intelectuales, entre otras (Cormier *et al.*, 2010). La definición de los grupos y subgrupos de alumnos con discapacidad y el diseño de acomodaciones que den respuesta a sus necesidades para promover el acceso a la prueba también presenta desafíos importantes. Por ejemplo, la elaboración de versiones especiales de las pruebas y materiales adicionales para alumnos que tienen una discapacidad o limitación visual, requiere que haya habido un diagnóstico previo, que se reconozca la diversidad de necesidades que tienen los sustentantes, se dimensionen, y se valore la factibilidad de su diseño y aplicación. En países como Inglaterra y los Estados Unidos de Norteamérica se apoya en diagnósticos previos y en el conocimiento de las necesidades de los niños, logrado a través del vínculo entre la escuela y la familia.

3. El uso de acomodaciones en las pruebas estandarizadas requiere de trabajos previos de investigación evaluativa (a nivel teórico y empírico) sobre cada una de las acomodaciones que se desea implementar. Estudios como el de Solano-Flores y sus colegas (2014) son ejemplo de lo que se debería realizar para evaluar el funcionamiento de una acomodación; éste dependerá al menos de tres factores: el tipo de acomodación que se desea implementar, el tópico que se desea evaluar y el tipo de necesidades que tienen los estudiantes.

Los estudios empíricos pueden considerar la inclusión de procesos de experimentación a fin de contar con más elementos de evaluación y para la toma de decisiones.

4. Es necesario que la implementación de las acomodaciones de una prueba esté acompañada de investigaciones sobre los efectos que tienen en los resultados de los alumnos que reciben la acomodación, y comparar éstos con los de quienes no recibieron acomodaciones.

Algunas investigaciones sobre la efectividad de las acomodaciones han seguido un método de análisis en el cual comparan el desempeño promedio en una prueba con ajustes, de los alumnos cuya primera lengua difiere de la lengua en que se aplica la prueba, y de aquellos que tuvieron la versión sin ajustes. Este método ha sido criticado, pues un aumento en el desempeño promedio podría tener que ver con que la prueba ajustada sea más sencilla, y no con que se haya reducido la varianza irrelevante del constructo. Se sugiere que se analicen los cambios en puntajes promedios para los grupos de alumnos cuya primera lengua es distinta a la lengua en que se aplica la prueba y que necesitan acomodaciones, y aquellos que no tienen necesidades relacionadas con el lenguaje de la prueba y que tampoco muestran alguna discapacidad (Pennock-Roman y Rivera, 2011).

5. Los procesos para la identificación de grupos y subgrupos con necesidades específicas a ser atendidas para ampliar el acceso a las pruebas estandarizadas y el diseño de acomodaciones que les den respuesta han sido impulsados en otros países por un marco legislativo. Por ejemplo, en los Estados Unidos de Norteamérica, la legislación federal estableció que las evaluaciones deberían hacerse disponibles con adaptaciones para los estudiantes con distintas necesidades de aprendizaje y también para aquellos alumnos cuya primera lengua fuera distinta a la lengua en que se aplicaba la prueba; asimismo, que los resultados deberían ser desagregados por género, grupo racial y étnico, y el estado de dominio de la lengua en que se aplicara la prueba. En nuestro país, un soporte legislativo como éste podría impulsar también la definición de procesos para identificar las necesidades de los alumnos, la exploración de diferentes alternativas para satisfacer estas necesidades y la investigación sobre los efectos de los ajustes en los resultados de los alumnos.
6. En la revisión de las prácticas en materia de consideración a la diversidad, se identifica un vínculo entre éstas y la manera en la cual se conceptualiza el derecho a la evaluación por parte de los alumnos. Por ejemplo, en los Estados Unidos de América del Norte e Inglaterra, se asume que todos los alumnos tienen derecho a ser evaluados, que las mediciones sean confiables y sus interpretaciones válidas. Esta concepción impulsa un enfoque hacia la maximización de la accesibilidad de las pruebas para los sustentantes, de tal manera que puedan demostrar sin obstáculos su desempeño.
7. Reducir las brechas en el desempeño de los alumnos que necesitan ajustes en las pruebas estandarizadas para facilitar su acceso a ellas, requiere no sólo de la definición de las acomodaciones pertinentes, sino de mejorar las oportunidades de aprendizaje de los alumnos. El estudio de Abedi *et al.* (2006) fue el primero en mostrar el impacto que tienen las oportunidades de aprender en el desempeño de los alumnos, independientemente de los esfuerzos para mejorar la validez de las pruebas a través de las acomodaciones.

Los estudios sobre la efectividad de las acomodaciones han mostrado también que ésta no puede ser valorada cuando los estudiantes no han tenido la oportunidad de aprender el material contenido en la prueba.

8. La elección de las acomodaciones deberá tener en cuenta las prácticas de enseñanza a las que han estado sujetos los estudiantes. Se ha encontrado que la implementación de una acomodación que desconocen o a la que no están acostumbrados los estudiantes, puede introducir varianza irrelevante al constructo y no contribuir a una medición válida de los aprendizajes.
9. Las medidas para maximizar la accesibilidad pueden ser de diferente naturaleza. Una de las que podría ser implementada en todos los sistemas de evaluación tiene que ver con el lenguaje utilizado para la redacción de los ítems y su presentación. Esto implica considerar principios de accesibilidad de lenguaje en el diseño de los ítems en lo individual y de la prueba en conjunto, relacionados con: las características lingüísticas (dificultad de las palabras y oraciones), cohesión y coherencia, estructura del contenido y complejidad, características de la impresión para facilitar su legibilidad, y organización de los textos (Janan y Wray, 2012).

10. La metodología del Diseño Universal de las Evaluaciones, descrita en el Marco Conceptual de este informe, podría ser una herramienta que contribuya a la consideración de la atención a la diversidad en todas las fases del proceso evaluativo.

A continuación se enuncian conclusiones específicas para cada uno de los sistemas evaluativos descritos a lo largo de este informe.

## SISTEMA DE EVALUACIÓN DE CHILE

Las pruebas del Sistema de Medición de la Calidad de la Educación (SIMCE) son un ejemplo notable en Latinoamérica sobre la consideración del contexto socioeconómico en las evaluaciones estandarizadas de logro. Contextualizar los resultados es un acierto porque permite hacer comparaciones más justas, al tiempo que posibilita el establecimiento de metas plausibles en el progreso del rendimiento escolar.

La consideración sobre el contexto socioeconómico ocurre en varias fases del proceso de evaluación, lo que constituye una práctica apropiada. Si bien las pruebas regulares del SIMCE han avanzado de manera notable en la consideración de la situación socioeconómica de los estudiantes, todavía hace falta mejorar algunos aspectos. Por ejemplo:

- En la fase de diseño, en los comités de validación de los reactivos, además de las figuras que actualmente se incluyen, sería deseable disponer de especialistas en las principales culturas que prevalecen en Chile, tales como antropólogos, lingüistas y sociólogos.
- Incluso es recomendable disponer de estos especialistas en las primeras revisiones de las especificaciones de los reactivos, para que, durante el análisis que se realiza para valorar la correspondencia entre dichas especificaciones y el currículo, se considere la diversidad socioeconómica o aquellas fuentes de diversidad que sean de especial interés.

La prueba del SIMCE para alumnos con Discapacidad Sensorial es una muestra positiva de cómo las evaluaciones externas del aprendizaje de los alumnos pueden adaptarse para ser más inclusivas y no discriminar a ciertas subpoblaciones.

## SISTEMA EDUCATIVO DE LOS ESTADOS UNIDOS DE NORTE AMÉRICA

La inclusión de los estudiantes con algún tipo de discapacidad en los procesos educativos implica realizar ciertos ajustes o acomodaciones que les permitan desempeñarse en igualdad de condiciones en el grupo escolar. Las evaluaciones de logro requieren, en muchos casos, de una disposición de infraestructura distinta a la de la enseñanza, y de manera particular en la aplicación de evaluaciones estandarizadas en gran escala. La revisión de la literatura muestra que hay grandes avances en la atención a la inclusión de los estudiantes con algún tipo de discapacidad en este tipo de evaluaciones. En los Estados Unidos y en el contexto de la ley No Child Left Behind (NCLB) existen definiciones claras de los servicios especiales o acomodaciones necesarios no sólo para el acceso a la enseñanza, sino también para la valoración en circunstancias igualitarias y sin que la discapacidad sea un factor de obstáculo cuando el estudiante es evaluado.

Se propone que la selección del tipo de acomodación o de servicio especial pertinente de acuerdo a las necesidades de los estudiantes sea el resultado de una reflexión profunda y del análisis de una serie de cuestiones sobre dos asuntos principales: la prueba en sí misma y la necesidad de acomodaciones, excluyendo aquellas que pudieran invalidar la evaluación.

Pero no sólo la selección del tipo de acomodación para los procesos de aula y para el momento de la evaluación es importante; la revisión de la efectividad de las acomodaciones en ambas situaciones se reconoce como una estrategia que permitirá reajustar los servicios especiales siguiendo el principio de no invalidar las evaluaciones.

Con relación a los estudiantes que son evaluados en una lengua distinta a su lengua materna, el número de estudiantes en los Estados Unidos que son considerados como English Language Learners (ELL) ha mostrado un crecimiento sustancial desde 1993 (Gottlich, 2006; en Solano y Gustafson, 2013). En el ciclo escolar 2005-2006 existían en los Estados Unidos más de 10.7 millones de estudiantes cuya lengua materna es distinta al inglés, constituyendo 20% de la población estudiantil (Kewal Ramani, Gilbertson, Fox y Provasnik, 2007; en Solano y Gustafson, 2013). Para 2002 había 4.5 millones de estudiantes clasificados con un desempeño limitado en inglés (LEP, por sus siglas en inglés), de los cuales 9.3% estaba representado por estudiantes de Pre-K hasta 12° grado en escuelas públicas.

Es evidente que son necesarias estrategias que permitan, por un lado, incluir a los estudiantes cuya lengua materna no es el inglés en los exámenes estandarizados de logro nacionales, al tiempo que no se comprometa la validez de los resultados, dado que las evaluaciones dependen en gran medida del uso de la lengua oficial del país. Las estrategias descritas muestran una preocupación por incluir a todos los estudiantes del sistema educativo, a la vez que reconocen la dificultad que significa esta inclusión en términos de la confiabilidad de los resultados de los estudiantes con LEP.

La adaptabilidad o traducción de los exámenes estandarizados, si bien trata de minimizar los resultados de logro no confiables, conlleva dificultades; por un lado de interpretación, regionalismos, vocabulario, etcétera, y por otro, puede crear confusión en los estudiantes, toda vez que la enseñanza de los contenidos que reciben en el salón de clases es en inglés y en los casos en los que las clases se imparten de manera bilingüe, es posible que subsistan problemas, por ejemplo de vocabulario. Solano Flores (2012) identifica cuatro acomodaciones de traducción como viables en las pruebas para los ELL: una versión de prueba en el idioma nativo, una versión bilingüe de la prueba, una versión de la prueba con instrucciones en el idioma nativo del estudiante y un glosario bilingüe. Solano-Flores (2012) señala que la traducción de exámenes es una tarea compleja que va más allá de la simple traducción y tendría que considerar la enorme heterogeneidad de la población de los ELL, las fallas potenciales de la acomodación de la traducción, y la necesidad de coordinar esfuerzos con los organismos y los colegas que son ajenos al proceso de traducción, pero cuyas acciones influyen en la integridad de las traducciones de prueba.

## **SISTEMA DE EVALUACIÓN INGLÉS**

Las pruebas estandarizadas de logro utilizadas en las etapas clave 1 y 2, reportadas en este informe, permitieron identificar las prácticas para maximizar el acceso a las pruebas a diversidad de alumnos sustentantes. Se reportaron prácticas de atención a alumnos con diferentes tipos de

discapacidad, y también con otras características que no son consideradas como discapacidad pero que pueden afectar la posibilidad de mostrar sus talentos en una prueba. Además de la diversidad en el manejo de la lengua inglesa, no se encontraron reportes de prácticas relacionadas con el contexto cultural del cual provienen los alumnos.

En el sistema educativo inglés, la decisión sobre las acomodaciones o modificaciones de acceso a emplear recae en gran parte en la escuela. Para esto, obviamente se requiere de la interacción de múltiples actores educativos: docentes con conocimiento de las características de los alumnos, directores al tanto de las necesidades de los alumnos, y padres interesados en que sus hijos sean evaluados bajo condiciones justas. Además, participa el sistema de salud con la dictaminación de las necesidades especiales que pueden tener los niños y los recursos que se otorgan para su aprendizaje de manera constante.

Los ajustes en las aplicaciones deben reflejar las prácticas cotidianas en la escuela, es decir, la manera en la cual los alumnos aprenden. Si regularmente utilizan lectores, traductores, apuntes, o dispositivos que ayuden a la lectura de textos o indicaciones, en las pruebas también se les debe dar acceso a estos medios.

Al revisar este sistema de evaluación no se encontraron estudios donde se emplearan técnicas de análisis del funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés), o se valorara la validez de los resultados que arrojan. Esta documentación es necesaria para sustentar que las adecuaciones no están afectando el constructo medido.

## SISTEMA DE EVALUACIÓN AUSTRALIANO

El sistema educativo australiano busca cerrar la brecha en los resultados de logro de los estudiantes aborígenes e isleños a través de un plan de acción bien definido que incluye una evaluación continua que retroalimenta las acciones de ese plan de acción. El sistema de evaluación educativo opera a nivel de sistema, de las escuelas, de los maestros y de los estudiantes. Así, en Australia se aplica el Programa de Evaluación Nacional en Alfabetización y Aritmética (NAPLAN, por sus siglas en inglés), en el que se advierten esfuerzos para atender a la diversidad de la población. Dentro de estos esfuerzos se contempla a las poblaciones indígenas mediante la inclusión de expertos en lenguas y cultura indígenas, aunque los expertos indígenas señalan no estar completamente seguros de que se logre el acceso a esta población a través de la prueba.

Las escuelas tienen la responsabilidad de reducir los efectos de la desventaja educativa asociada al estatus socioeconómico, la condición indígena y la discapacidad. La equidad se define de una manera horizontal, cuando los servicios son igualmente accesibles a todos los miembros en la comunidad con niveles similares de necesidades, y otro vertical, cuando los servicios representan las necesidades especiales de determinados grupos de la comunidad y se ajustan aspectos de la prestación de servicios para satisfacer estas necesidades.

La atención a la diversidad también incluye acomodaciones para niños con necesidades educativas especiales. Al igual que en Inglaterra, las acomodaciones se definen en función de los ajustes que se hacen de manera cotidiana en la enseñanza; sin embargo, no parece haber consistencia en su uso durante las evaluaciones.



## PROGRAMA PARA LA EVALUACIÓN INTERNACIONAL DE ESTUDIANTES

Esta prueba sigue reglas de exclusión cuidando la representatividad de los países; existen varias razones por las que una escuela o un estudiante pueden ser excluidos de tomar la prueba del Programa para la Evaluación Internacional de Estudiantes (PISA, por sus siglas en inglés); las escuelas pueden ser excluidas porque se encuentran localizadas en regiones remotas y de difícil acceso, porque son muy pequeñas, o debido a factores organizativos u operacionales que imposibilitan la participación. Otras razones de exclusión incluyen una discapacidad intelectual o limitaciones de su desempeño en el idioma de la evaluación. No obstante, la atención a la diversidad en la aplicación de esta prueba estandarizada se observa en relación con el contexto lingüístico, variable a la que se le pone especial atención en la búsqueda de equivalencias que permitan asegurar que lo que se pregunta o plantea en un idioma sea un reflejo en espejo de lo que se pregunta o plantea en otra lengua. Es claro que existen diferencias culturales que van más allá de la lengua y que plasmarlas o reflejarlas en un instrumento de evaluación no es un trabajo menor.

Las propuestas para lograr una equivalencia adecuada son pertinentes; la traducción paralela y la conciliación parecen contribuir a desarrollar una versión adecuada en la lengua objetivo.

## SISTEMA DE MONITOREO OREALC

El Sistema Regional de Información Educativa de los Estudiantes con Discapacidad (SIRIED) que desarrolla la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC) es un ejemplo de sistemas de monitoreo de calidad educativa con enfoque de atención a la diversidad. Bajo la posición de que lo que se evalúa adquiere importancia, la información que recoge el SIRIED permite dar una mayor visibilidad a las personas con discapacidad. El INEE podría utilizar estrategias similares para promover una educación inclusiva que considere paulatinamente a grupos de especial interés.

El proceso de definición de los sistemas de monitoreo, al igual que como ocurre con el SIRIED, debería responder a múltiples usuarios de la información, lo que implica diferentes niveles de desglose. También convendría que el nivel de desagregación de los indicadores incluyera, como en el SIRIED, subpoblaciones en las que se desea asegurar una educación de calidad, por ejemplo, desagregar por género, tamaño de la localidad donde habitan las personas, nivel socioeconómico de las familias a las que pertenecen los subgrupos de interés, entre otros.

Del proceso de diseño y definición del sistema hay varios elementos de aprendizaje:

- La colaboración cercana con los usuarios permitió que los indicadores fueran más relevantes y pertinentes, lo que a su vez generó más conciencia sobre la importancia de disponer de sistemas de información como el SIRIED.
- Los piloteos en instituciones educativas permitieron refinar los indicadores y valorar la factibilidad de su construcción.

Los sistemas de información con un enfoque de inclusión, requieren la generación de información que no existe, o bien, no está sistematizada. Aunque no se deberían ignorar los recursos necesarios para adquirir la información con la que se generarán los indicadores, se debería

priorizar su utilidad para monitorear el grado en que el derecho a una educación de calidad está siendo efectivo para toda la población de interés.

## SISTEMA DE EVALUACIÓN SACMEQ

En el contexto africano, el Consorcio para Monitorear la Calidad Educativa en el Sur y Este de África (SACMEQ, por sus siglas en inglés) como sistema de monitoreo de la calidad de la educación, atiende a una problemática sensible en aquella región: la desigualdad educativa entre hombres y mujeres. El SACMEQ utiliza como línea de atención a la diversidad la cuestión del sexo, no sólo de estudiantes sino también de los docentes y directivos.

Aunque el SACMEQ está al pendiente de la atención a la diversidad del género en varias de las etapas del proceso de evaluación de los aprendizajes y monitoreo de la calidad educativa, aún falta que dicho sistema intensifique el énfasis de inclusión de la mujer en las primeras etapas de los procesos evaluativos, tales como el diseño de los instrumentos y en los análisis psicométricos.

# Referencias bibliográficas

- ACARA. (2014a). 2014 NAPLAN Handbook for principals. Australian Curriculum, Assessment and Reporting Authority.
- ACARA. (2014b). 2014 NAPLAN Test administration handbook for teachers. Year 3. Australian Curriculum, Assessment and Reporting Authority.
- ACARA. (2014c). 2014 NAPLAN Test administration handbook for teachers. Years 3 & 5. Australian Curriculum, Assessment and Reporting Authority.
- ACARA. (2014d). 2014 NAPLAN Test administration handbook for teachers. Years 7 & 9. Australian Curriculum, Assessment and Reporting Authority.
- AERA, APA y NCME. (2014). *Standards for educational and psychological testing*. Washington, D.C.: AERA/ APA/NCME.
- Agencia de Calidad de la Educación. (2013a). *Informe de Resultados SIMCE Discapacidad Sensorial 4° Básico 2012 para Docentes y Directivos*. Santiago, Chile: Agencia de Calidad de la Educación.
- Agencia de Calidad de la Educación. (2013b). *Informe Técnico SIMCE 2012*. Santiago, Chile: Agencia de Calidad de la Educación.
- Agencia de Calidad de la Educación. (2013c). Metodología de Construcción de Grupos Socioeconómicos SIMCE 2013. Agencia de Calidad de la Educación.
- Agencia de Calidad de la Educación. (2014). Calendario de Evaluaciones de la Agencia de Calidad de la Educación. Recuperado a partir de [https://s3-us-west-2.amazonaws.com/documentos-web/calendario\\_evaluaciones\\_Agencia.pdf](https://s3-us-west-2.amazonaws.com/documentos-web/calendario_evaluaciones_Agencia.pdf)
- Amadhila, L., Dengeinge, R., Miranda, H. y Shikongo, S. (2011). Progress in Gender Equality in Education: Namibia. SACMEQ *Policy Brief* (6).
- Atorresi, A., Pardo, C., Glejberman, D., Espinosa, G., Toranzos, L., Rocha, M. y Castro, M. (2010). *Compendio de los Manuales del Serce*. Santiago, Chile: OREALC/UNESCO. Recuperado a partir de <http://unesdoc.unesco.org/images/0019/001919/191940s.pdf>
- August, D. y McArthur, E. (1996). *Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of*. DIANE Publishing. Recuperado a partir de [http://books.google.com/books?hl=en&lr=&id=RpjLoRQ7m\\_EC&oi=fnd&pg=PR3&dq=%22document+summarizes+the+results+of+a+working+meeting+held+in+Washington,+DC%22+%22of+alternative+testing+procedures,+and+keeping+the+decision+flow+simple,+consistent,%22+%22than+years+in+English-only+instruction+or+other+background+characteristics.+This+is%22+&ots=1AK49GB5PO&sig=dWxKbdD260Bu-pqQo327fb6OnWE](http://books.google.com/books?hl=en&lr=&id=RpjLoRQ7m_EC&oi=fnd&pg=PR3&dq=%22document+summarizes+the+results+of+a+working+meeting+held+in+Washington,+DC%22+%22of+alternative+testing+procedures,+and+keeping+the+decision+flow+simple,+consistent,%22+%22than+years+in+English-only+instruction+or+other+background+characteristics.+This+is%22+&ots=1AK49GB5PO&sig=dWxKbdD260Bu-pqQo327fb6OnWE)
- Bew, P. (2011). Independent Review of Key Stage 2 testing, assessment and accountability. Final Report. Crown. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/176180/Review-KS2-Testing\\_final-report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf)
- Blanco, R., Mascardi, L. y Narvarte, L. (2010). *Sistema Regional de Información Educativa de los Estudiantes con Discapacidad*. SIRIED. *Propuesta metodológica*. Santiago, Chile: OREALC/UNESCO.
- Branch, J. L. (2001). Junior High Students and Think Alouds: Generating Information-Seeking Process Data Using Concurrent Verbal Protocols. *Library & Information Science Research*, 23(2), 107-122.
- British Dyslexia Association. (2015). Access arrangements. Recuperado 23 de enero de 2015, a partir de <http://www.bdadyslexia.org.uk/educator/access-arrangements>
- Casassus, J., Froemel, J., Palafox, J., y Cusato, S. (1998). *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado*. Santiago, Chile: LLECE-OREALC-UNESCO.
- Centro de Estudios MINEDUC. (2013). *Estadísticas de la Educación 2013*. Santiago, Chile: Ministerio de Educación. Chile.

- Chitiga, Z. M., y Chinoona, A. (2011). Progress in Gender Equality in Education: Zimbabwe. *SACMEQ Policy Brief*, (6).
- City of York Council. (2015). Educating Your Child at Home. Recuperado 19 de enero de 2015, a partir de [http://www.york.gov.uk/info/200334/school\\_services/294/educating\\_your\\_child\\_at\\_home](http://www.york.gov.uk/info/200334/school_services/294/educating_your_child_at_home)
- Cortiella, C. (2005). No Child Left Behind: Determining Appropriate Assessment Accommodations for Students with Disabilities. *National Center for Learning Disabilities*. Recuperado a partir de <http://eric.ed.gov/?id=ED486451>
- Davey, L. (1992). The Case for a National Testing System. ERIC Digest. Recuperado a partir de <http://eric.ed.gov/?id=ED410239>
- Department for Education. (2014a). Assessment, Curriculum and Qualifications. Research priorities and questions. The Department for Education. Recuperado a partir de [www.gov.uk/government/publications](http://www.gov.uk/government/publications)
- Department for Education. (2014b). Performance - P Scale - attainment targets for pupils with special educational needs. Department for Education.
- Department for Education. (2014c). Statistical First Release. National Curriculum Assessments at Key Stage 2 in England, 2014 (Provisional). Department for Education. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/347653/SFR30\\_2014\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/347653/SFR30_2014_Text.pdf)
- Department for Education and Skills. (2001). Special Educational Needs Code of Practice. Department for Education and Skills.
- Dolata, S. (2011). *Gender Equality in Education: Looking Beyond Parity*. Paris: IIEP-UNESCO.
- Douglas Silas Solicitors. (2015). Independent and Non-Maintained Special Schools. Recuperado 17 de enero de 2015, a partir de <http://www.specialeducationalneeds.co.uk/independent-and-non-maintained-special-schools.html>
- Eroles, D., y Hirmas, C. (2009). *Experiencias educativas de segunda oportunidad: Lecciones desde la práctica innovadora en América Latina*. Santiago, Chile: UNESCO.
- ETS. (2009). ETS Guidelines for Fairness Review of Assessments. Educational Testing Service. Recuperado a partir de [https://www.ets.org/Media/About\\_ETS/pdf/overview.pdf](https://www.ets.org/Media/About_ETS/pdf/overview.pdf)
- ETS. (2014a). Bulletin Supplement for Test Takers with Disabilities or Health-Related Needs. ETS. Recuperado a partir de [http://www.ets.org/s/gre/pdf/bulletin\\_supplement\\_test\\_takers\\_with\\_disabilities\\_health\\_needs.pdf](http://www.ets.org/s/gre/pdf/bulletin_supplement_test_takers_with_disabilities_health_needs.pdf)
- ETS. (2014b). Common Testing Accommodations. ETS. Recuperado a partir de [https://www.ets.org/disabilities/test\\_takers/accommodations/](https://www.ets.org/disabilities/test_takers/accommodations/)
- Francis, D.J., Rivera, M., Lesaux, N., Kieffer, M., y Rivera, H. (2006). Research-Based Recommendations for the Use of Accommodations in Large-Scale Assessments. Practical Guidelines for the Education of English Language Learners. Book 3 of 3. *Center on Instruction*. Recuperado a partir de <http://eric.ed.gov/?id=ED517792>
- GACE. (2014). Bulletin Supplement of Test Takers with Disabilities of Health-Related Needs. Taylor & Francis: Routledge.
- Gee, J. P. (2003). Opportunity to Learn: A Language-Based Perspective on Assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27-46. <http://doi.org/10.1080/09695940301696>
- Gipps, C., y Stobart, G. (2010). Fairness. En *International Encyclopedia of Education* (3.ª ed., Vol. 4, pp. 56-60). Oxford, UK.: Elsevier.
- Government Digital Service. (2014). Home Education. Recuperado 17 de enero de 2015, a partir de <https://www.gov.uk/home-education>
- Government Digital Service. (2015). Types of Schools. Recuperado a partir de <https://www.gov.uk/types-of-school>
- He, Q., Anwyll, S., Glanville, M., y Opposs, D. (2014). An Investigation of Measurement Invariance of the Key Stage 2 National Curriculum Science Sampling Test In England. *Research Papers in Education*, 29(2), 211-239.

- Hevia, R., Hirmas Ready, C., y Unesco. (2005). *Políticas educativas de atención a la diversidad cultural: Brasil, Chile, Colombia, México y Perú*. Santiago: UNESCO.
- Hirmas, R., y Carolina, B., Rosa. (2008). *Educación y diversidad cultural: lecciones desde la práctica innovadora en América Latina*. Santiago, Chile: UNESCO [u.a.].
- Hungi, N. (2010). Is There Gender Equity in School Managerial Positions?, *SACMEQ Policy Issues Series* (8).
- Hungi, N. (2011a). *Accounting for Variations in the Quality of Primary School Education*. París: SACMEQ-IIEP. Recuperado a partir de [http://www.sacmeq.org/sites/default/%20files/sacmeq/reports/sacmeq-iii/working-papers/07\\_multivariate\\_final.pdf](http://www.sacmeq.org/sites/default/%20files/sacmeq/reports/sacmeq-iii/working-papers/07_multivariate_final.pdf)
- Hungi, N. (2011b). *Characteristics of Grade 6 pupils, their homes, and learning environments*. Paris: SACMEQ-IIEP.
- Hungi, N. (2011c). *Characteristics of School Heads and their Schools*. Paris: SACMEQ-IIEP.
- Janan, D., y Wray, D. (2012). Guidance on the Principles of Language Accessibility in National Curriculum Assessments. Referring to the Regulatory Arrangements for National Assessments: National Curriculum and Early Year Foundation Stage. OFQUAL. Recuperado a partir de [http://dera.ioe.ac.uk/15879/1/2012-10-12-guidance-principles-of-language-accessibility-incas.pdf\\_dm\\_i%3DBTP,11764,301B94,35WJY,1](http://dera.ioe.ac.uk/15879/1/2012-10-12-guidance-principles-of-language-accessibility-incas.pdf_dm_i%3DBTP,11764,301B94,35WJY,1)
- Johnstone, C., Altman, J., y Thurlow, M. (2006). A State Guide to the Development of Universally Designed Assessments. *National Center on Educational Outcomes, University of Minnesota*. Recuperado a partir de <http://eric.ed.gov/?id=ED495885>
- Johnstone, C.J., Thompson, S.J., Bottsford-Miller, N.A., y Thurlow, M.L. (2008). Universal Design and Multimethod Approaches to Item Review. *Educational Measurement: Issues and Practice*, 27(1), 25-36.
- Jopo, H., Maema, M., y Ramokoena, M. (2011a). Progress in Gender Equality in Education: Lesotho. *SACMEQ Policy Brief* (6).
- Jopo, H., Maema, M., y Ramokoena, M. (2011b). The SACMEQ III Project in Lesotho. SACMEQ.
- Kaba, A., y Musonda, B. (2011). Progress in Gender Equality in Education: Zambia. *SACMEQ Policy Brief*, (6).
- Kupper, L. (2000). A Guide to the Individualized Education Program. Recuperado a partir de <http://eric.ed.gov/?id=ED444279>
- Leiva, P., Saavedra, X., y Muñoz, M. (2013). *Informe de análisis de Sistema Educativo con análisis de género. 2012*. Santiago, Chile: Ministerio de Educación. Chile.
- Leste, A., y Benstrong, E. (2010). The SACMEQ III Project in Seychelles. SACMEQ.
- Leste, A., y Benstrong, E. (2011). Progress in Gender Equality in Education: Seychelles. *SACMEQ Policy Brief*, (6).
- LLECE. (2013). TERCE. *Análisis curricular*. LLECE-OREALC-UNESCO.
- Makuwa, D. (2011). *Characteristics of Grade 6 Teachers*. París: SACMEQ-IIEP.
- Manzi, J., González, R., y Sun, Y. (Eds.). (2011). *La Evaluación Docente en Chile*. Santiago, Chile: MIDE-UC.
- Martín, M.T.O.B.S., y Alessandri, A.J. P. (2010). Profiles of Mathematics Performance by Students of Chilean City Schools, Based on Cognitive Ability by Socioeconomic Level. *Revista Electrónica de Investigación Educativa*, 12(1), 2.
- Matear, A. (2007). Equity in Education in Chile: The Tensions Between Policy and Practice. *International Journal of Educational Development*, 27(1), 101-113. <http://doi.org/10.1016/j.ijedudev.2006.06.015>
- Mhonyiwa, J.E., Ponera, G. y Mrutu, A.S. (2011a). Progress in Gender Equality in Education: Tanzania Mainland. *SACMEQ Policy Brief*, (6).
- Mhonyiwa, J.E., Ponera, G. y Mrutu, A.S. (2011b). The SACMEQ III Project in Tanzania. SACMEQ.
- Milner, G., Mulera, D., Banda, T.C., Matale, E., y Chimombo, J. (2011). Progress in Gender Equality in Education: Malawi. *SACMEQ Policy Brief*, (6).
- Milner G., Mulera D. y Chimuzu T. (2011). The SACMEQ III Project in Malawi. SACMEQ.
- MINEDUC. (2013). Educación en una mirada. Año 2013. Recuperado a partir de [http://sgdce.mineduc.cl/descargar.php?id\\_doc=201403141245080](http://sgdce.mineduc.cl/descargar.php?id_doc=201403141245080)
- Miranda, H., Amadhila, L., Dengeinge, R., y Shikongo, S. (2011). The SACMEQ III Project in Namibia. SACMEQ.
- Mohammed, K.A., Mohamed, M., Abdalla, M., Masoud, K. y Sharif, O. (2010). The SACMEQ III Project in Zanzibar. SACMEQ.

- Moloi, M.Q., y Chetty, M. (2010). The SACMEQ III Project in South Africa. SACMEQ.
- Moloi, M.Q., y Chetty, M. (2011). Progress in Gender Equality in Education: South Africa. SACMEQ *Policy Brief*, (6).
- Monyaku, B., y Mmereki, O.A. (2010). The SACMEQ III Project in Botswana. SACMEQ.
- Monyaku, B., y Mmereki, O.A. (2011). Progress in Gender Equality in Education: Botswana. SACMEQ *Policy Brief* (6).
- Murimba, S. (2005). The Impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). *Prospects*, 35(1), 91-108.
- Musonda, B., y Kaba, A. (2010). The SACMEQ III Project in Zambia. SACMEQ.
- National Foundation for Educational Research. (2015). Information about Assessment. Key Stage 2 National Curriculum English Test. Recuperado 27 de enero de 2015, a partir de <http://www.nfer.ac.uk/research/centre-for-assessment/key-stage-2-national-curriculum-based-test.cfm>
- National Project Managers' Meeting. (2010a). Material Preparation Field Trial 2012. OECD.
- National Project Managers' Meeting. (2010b). Translation and Adaptation Guidelines for PISA 2012. OECD.
- OECD. (2006). PISA 2006. Marco de la evaluación. OECD Publishing.
- OECD. (2012). PISA 2009 Technical Report. <http://doi.org/http://dx.doi.org/10.1787/9789264167872-en>
- OECD. (2013). *PISA 2012 Results: Ready to Learn Students' Engagement, Drive and Self-Beliefs* (Vol. II). Paris: OECD Publishing. Recuperado a partir de <http://dx.doi.org/10.1787/9789264201170-en>
- OECD. (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.
- OECD. (s. f.). *El programa PISA de la OCDE. Qué es y para qué sirve*. Paris: OECD Publishing. Recuperado a partir de <http://www.oecd.org/pisa/39730818.pdf>
- Office of Learning/Instruction, Standards, Assessment, and Accountability Unit. (2014). Oregon Accessibility Manual. Oregon Department of Education. Recuperado a partir de <http://www.ode.state.or.us/search/page/?=487>
- Office of Qualifications and Examinations Regulation. (2011a). Guidance on Monitoring Access to National Curriculum Assessments. Regulatory Arrangements for National Assessments: National Curriculum and Early Years Foundation Stage. OFQUAL. Recuperado a partir de [http://dera.ioe.ac.uk/15878/2/2012-10-12-guidance-monitoring-access-to-national-curriculum-assessments\\_Redacted.pdf](http://dera.ioe.ac.uk/15878/2/2012-10-12-guidance-monitoring-access-to-national-curriculum-assessments_Redacted.pdf)
- Office of Qualifications and Examinations Regulation. (2011b). Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage. OFQUAL. Recuperado a partir de <http://dera.ioe.ac.uk/10033/1/2011-regulatory-framework-for-national-assessments.pdf>
- Oklahoma State Department of Education. (s. f.). OSTP Accommodations for Student with an Individualized Education Program (IEP) or Section 504 Plan. Oklahoma State Department of Education.
- Olson, J.F., y Goldstein, A.A. (1997). *The Inclusion of Students with Disabilities and Limited English Proficient Students in Large-Scale Assessments. A Summary Of Recent Progress*. DIANE Publishing. Recuperado a partir de [http://books.google.com/books?hl=en&lr=&id=ajpT38xfe00C&oi=fnd&pg=PR2&dq=%22report+is+one+of+a+continuing+series+of+steps+that+indicate+our+commitment+and+progress+to+ward%22+%22Director+for+Item+Development+and+Operations+in+the+Assessment+Group+at+N CES,+and%22+&ots=\\_SLkSH1Yb4&sig=LhXYL0HvpvGldVSfD3bk5ORqNaU](http://books.google.com/books?hl=en&lr=&id=ajpT38xfe00C&oi=fnd&pg=PR2&dq=%22report+is+one+of+a+continuing+series+of+steps+that+indicate+our+commitment+and+progress+to+ward%22+%22Director+for+Item+Development+and+Operations+in+the+Assessment+Group+at+N CES,+and%22+&ots=_SLkSH1Yb4&sig=LhXYL0HvpvGldVSfD3bk5ORqNaU)
- OREALC. (2012). SIRIED. Informe final. OREALC /UNESCO.
- OREALC. (2013). *SIRIED. Resultados de la primera fase de aplicación*. Santiago, Chile: OREALC /UNESCO.
- Passos, A., Nahara, T., y Magaia, F. (2012). Progress in Gender Equality in Education: Mozambique. SACMEQ *Policy Brief* (6).
- Pennock-Roman, M., y Rivera, C. (2011). Mean Effects of Test Accommodations for ELLs and Non-ELLs: A Meta-Analysis of Experimental Studies. *Educational Measurement: Issues and Practice*, 30(3), 10-28.
- QCDA. (2010). Test Development Level Setting and Maintaining Standards. Qualifications and Curriculum Development Agency. Recuperado a partir de <http://dera.ioe.ac.uk/12005/1/assessments%20test%20setting.pdf>

- Rico, L. (2006). Marco teórico de evaluación en PISA sobre matemáticas y resolución de problemas. *Revista de Educación*, (Extraordinario), 275-294.
- Sabatini, J.P., y Bruce, K.M. (2011). *PISA Reading Components Translation and Adaptation Guidelines*. Princeton, NJ, USA: Educational Testing Service.
- SACMEQ. (1995a). Manual for Data Collectors. SACMEQ-IIEP.
- SACMEQ. (1995b). Manual for National Research Co-Ordinators. SACMEQ-IIEP.
- SACMEQ. (2008). Data Building and Data Management. SACMEQ-IIEP.
- Saito, M. (2004). Gender Equality in Reading and Mathematics Achievement. *IIEP Newsletter*, (April-June).
- Saito, M. (2010a). Has Gender Equality in Reading and Mathematics Achievement Improved? *SACMEQ Policy Issues Series*, (4). Recuperado a partir de [www.sacmeq.org](http://www.sacmeq.org)
- Saito, M. (2010b). The gender gap in learning back to square one. *IIEP Newsletter*, 28(3), 5-6.
- Saito, M. (2011). *Trends in Gender Equality in Learning Achievement in Southern and Eastern Africa: Exploration of Characteristics of Educational Environment and Curriculum Areas*. París: IIEP-UNESCO. Recuperado a partir de [http://doc.iiep.unesco.org/wwwisis/repdoc/sem313/sem313\\_5\\_eng.pdf](http://doc.iiep.unesco.org/wwwisis/repdoc/sem313/sem313_5_eng.pdf)
- Saito, M., Ssenbulya, F., y Lubega, I. (2011). Progress in Gender Equality in Education: Uganda. *SACMEQ Policy Brief* (6).
- Santiago, P., Donaldson, G., Herman, J., y Shewbridge, C. (2011). OECD Reviews of Evaluation and Assessment in Education: Australia. *OECD Publishing (NJ1)*. Recuperado a partir de <http://eric.ed.gov/?id=ED528253>
- Sauba, D., y Lutchimiah, B. (2010). The SACMEQ III Project in Mauritius. SACMEQ.
- Sauba, D., y Lutchimiah, B. (2011). Progress in Gender Equality in Education: Mauritius. *SACMEQ Policy Brief* (6).
- Shabalala, J.T., y Nxumalo, P. (2011). Progress in Gender Equality in Education: Swaziland. *SACMEQ Policy Brief* (6).
- Shafer Willner, L., Rivera, C., y Acosta, B. D. (2008). Descriptive Study of State Assessment Policies for Accommodating English Language Learners. The George Washington University Center for Equity and Excellence in Education. Recuperado a partir de <http://www.ode.state.or.us/wma/teachlearn/common-core/descriptivestudyell.pdf>
- SIRNEE. (2007). *Consulta a países de Latinoamérica sobre información asociada a las NEE*. Santiago, Chile: OREALC/UNESCO.
- Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices. En *Cultural validity in assessment* (pp. 3-21). New York: Routledge.
- Solano-Flores, G. (2012a). Smarter Balanced Assessment Consortium: Translation Accommodations Framework for Testing English Language Learners in Mathematics. Smarter Balanced Assessment Consortium. Recuperado a partir de <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/Translation-Accommodations-Framework-for-Testing-ELL-Math.pdf>
- Solano-Flores, G. (2012b). Translation Accommodations Framework for Testing English Language Learners in Mathematics. Smarter Balanced Assessment Consortium.
- Solano-Flores, G., Backhoff, E., y Contreras Niño, L.Á. (2009). Theory of Test Translation Error. *International Journal of Testing*, 9(2), 78-91. <http://doi.org/10.1080/15305050902880835>
- Solano-Flores, G., Contreras Niño, L.Á., y Backhoff, E. (2006). Traducción y adaptación de pruebas: Lecciones aprendidas y recomendaciones para países participantes en TIMSS, PISA y otras comparaciones internacionales. *REDIE: Revista Electrónica de Investigación Educativa*, 8(2), 3.
- Solano-Flores, G., y Nelson-Barber, S. (2001). On the Cultural Validity of Science Assessments. *Journal of Research in Science Teaching*, 38(5), 553-573.
- Solano-Flores, G., Shade, C., y Chrzanowski, A. (2014). Item Accessibility and Language Variation. Conceptual Framework. Smarter Balanced Assessment Consortium.
- Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., y Nguyen-Le, K. (2014). Developing Testing Accommodations for English Language Learners: Illustrations as Visual Supports for Item Accessibility. *Educational Assessment*, 19(4), 267-283. <http://doi.org/10.1080/10627197.2014.964116>



- STA. (2014a). About Standards and Testing Agency. Recuperado a partir de <https://www.gov.uk/government/organisations/standards-and-testing-agency/about>
- STA. (2014b). Braille Tests for Children with a Visual Impairment. Modified Test Administrators' Guide. Key Stage 2. Standards & Testing Agency.
- STA. (2014c). Modified Large Print (MLP) for children with a visual impairment. Modified Test Administrators' Guide. Key Stage 2. Standards & Testing Agency.
- STA. (2014d). Tests for Children with a Hearing Impairment and Children Who Use Sign Language. Modified Test Administrators' Guide. Key Stage 2. Standards & Testing Agency.
- Standards & Testing Agency. (2014a). Assessment and Reporting Arrangements Key Stage 2. STA. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/278150/2014\\_KS2\\_assessmentandreportingarrangementsARA\\_DIGITAL\\_HO.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/278150/2014_KS2_assessmentandreportingarrangementsARA_DIGITAL_HO.pdf)
- Standards & Testing Agency. (2014). Key Stage 2 Mathematics Test Framework (draft). National Curriculum Tests from 2016. STA. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/295174/2016\\_Key\\_stage\\_2\\_Mathematics\\_test\\_framework.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295174/2016_Key_stage_2_Mathematics_test_framework.pdf)
- Standards & Testing Agency. (2014b). Key stage 2 Tests: How to Use Access Arrangements. STA. Recuperado a partir de <https://www.gov.uk/key-stage-2-tests-how-to-use-access-arrangements>
- Standards & Testing Agency. (2014c). National Curriculum Assessments. Key Stage 2 Dates and Deadlines. STA. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/350688/2015\\_KS2\\_wall\\_chart.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/350688/2015_KS2_wall_chart.pdf)
- Standards & Testing Agency. (2014d). Notes for Readers in the English Grammar, Punctuation and Spelling Sest: Short Answer Questions. Key Stage 2. April 2014. STA. Recuperado a partir de [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/296344/2014\\_Notes\\_for\\_readers\\_in\\_English\\_GPS\\_test.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/296344/2014_Notes_for_readers_in_English_GPS_test.pdf)
- Standards & Testing Agency. (2014e). Standards and Testing Agency Business Plan. 1 April 2014 - March 2015. STA.
- Stobart, G. (2005). Fairness in Multicultural Assessment Systems. *Assessment in Education: Principles, Policy & Practice*, 12(3), 275-287. <http://doi.org/10.1080/09695940500337249>
- Stobart, G. (2009). Determining Validity in National Curriculum Assessments. *Educational Research*, 51(2), 161-179. <http://doi.org/10.1080/00131880902891305>
- Thompson, S., Johnston, C.J., y Thurlow, M.L. (2002). *Universal Design Applied to Large Scale Assessments*. National Center on Educational Outcomes Synthesis Report. Recuperado a partir de [https://osepideasthatwork.org/Toolkit/pdf/Universal\\_Design\\_LSA.pdf](https://osepideasthatwork.org/Toolkit/pdf/Universal_Design_LSA.pdf)
- Treviño, E., Castro, M., Costilla, R., y Acevedo, C. (2008). *SERCE. Los aprendizajes de los estudiantes de América Latina y el Caribe. Primeros resultados*. (H. Valdés, ed.). Santiago, Chile: LLECE-OREALC-UNESCO.
- Treviño, E., Castro, M., Costilla, R., y Acevedo, C. (2009). *SERCE. Los aprendizajes de los estudiantes de América Latina y el Caribe. Reporte técnico*. (H. Valdés, ed.). Santiago, Chile: LLECE-OREALC-UNESCO.
- Unidad de Currículum y Evaluación. (2007). *SIMCE. Informe de Resultados Nacionales 2006*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2008). *SIMCE. Informe de Resultados Nacionales 2007*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2009). *SIMCE. Informe de Resultados Nacionales 2008*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2010). *SIMCE. Informe de Resultados Nacionales 2009*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2011a). *Informe de Resultados Estudiantes con Discapacidad Sensorial SIMCE 4º Básico 2010*. Santiago, Chile: Ministerio de Educación.



- Unidad de Currículum y Evaluación. (2011b). *SIMCE. Informe de Resultados Nacionales 2010*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2012a). *Informe de Resultados Estudiantes con Discapacidad Sensorial SIMCE 4° Básico 2011*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2012b). SIMCE 2011. Cuestionario Docentes. 4° Básico.
- Unidad de Currículum y Evaluación. (2012c). SIMCE 2011. Cuestionario Estudiantes. 4° Básico.
- Unidad de Currículum y Evaluación. (2012d). SIMCE 2011. Cuestionario Estudiantes. 8° Básico.
- Unidad de Currículum y Evaluación. (2012e). SIMCE 2011. Cuestionario Padres. 4° Básico.
- Unidad de Currículum y Evaluación. (2012f). SIMCE 2011. Cuestionario Padres. 8° Básico.
- Unidad de Currículum y Evaluación. (2012g). *SIMCE. Informe de Resultados Nacionales 2011*. Santiago, Chile: Ministerio de Educación.
- Unidad de Currículum y Evaluación. (2013). Otros Indicadores de Calidad Educativa. Fundamentos.
- Unidad de Currículum y Evaluación. (2014a). *Estándares Indicativos de Desempeño para los Establecimientos Educativos y sus Sostenedores*. Santiago, Chile: Ministerio de Educación. Chile.
- Unidad de Currículum y Evaluación. (2014b). *Otros Indicadores de Calidad Educativa*. Santiago, Chile: Ministerio de Educación. Chile.
- Varas, R., Guzmán, I., Ibarra, C., Peña, C., Rozas, C., Duk, C., y Godoy, P. (2003). *Cada escuela es un mundo un mundo de diversidad*. Santiago, Chile: OREALC/UNESCO-HINENI.
- Wasanga, P.M., Ogle, M.A., y Wambua, R.M. (2011). Progress in Gender Equality in Education: Kenya. *SACMEQ Policy Brief (6)*.
- Wasanga, P.M., Ogle, M.A., y Wambua, R.M. (2012). The SACMEQ III Project in Kenya. SACMEQ-Kenya National Examinations Council.
- Weston, T.J. (2002). The Validity of Oral Accommodation In Testing. *NAEP Validity Studies (NVS) Panel*. Recuperado a partir de [http://www.air.org/sites/default/files/downloads/report/weston\\_finalrevpdf\\_0.pdf](http://www.air.org/sites/default/files/downloads/report/weston_finalrevpdf_0.pdf)
- Zieky, M. (2006). Fairness review in assessment. En S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359-376). Mahwah, N.J.: L. Erlbaum.



ATENCIÓN A LA DIVERSIDAD  
EN EVALUACIONES EDUCATIVAS EXTERNAS.  
MUESTRA DE PRÁCTICAS INTERNACIONALES

Se terminó de imprimir en diciembre de 2015 en los talleres  
de Impresora y Encuadernadora Progreso S.A. de C.V. (IEPSA).

En su formación se utilizaron las familias  
tipográficas: Frutiger Lt Std y Museo.

Tiraje: 500 ejemplares



**Este documento responde a la solicitud expresa** del Instituto Nacional para la Evaluación de la Educación (INEE) de desarrollar un estudio que permitiera caracterizar, clasificar y analizar las distintas prácticas internacionales en materia de consideración de la diversidad en evaluaciones educativas externas.

El proceso de atención a la diversidad en evaluaciones estandarizadas es sumamente complejo, pues requiere identificar grupos de estudiantes con necesidades específicas, planificar la atención a estas necesidades, implementar medidas destinadas a ampliar el acceso a las pruebas y revisar constantemente la efectividad de tales medidas. Es por ello que no existen prácticas únicas para atender múltiples tipos de diversidad en las distintas fases del proceso evaluativo.

En ese sentido, las prácticas en materia de consideración a la diversidad de los países y organismos internacionales que se incluyen en este informe no deben verse como un modelo a seguir, pues obedecen a procesos y condiciones específicas que no coinciden necesariamente con los del Sistema Educativo Nacional. Sin embargo, dichas prácticas aportan ejemplos valiosos que enriquecen los procesos evaluativos de nuestro país, y contribuyen a avanzar en la justicia de las evaluaciones y a mejorar las oportunidades de aprendizaje de todos los alumnos.



Descargue una copia digital gratuita



Comuníquese con nosotros



Visite nuestro portal

