

# Experiencias Internacionales sobre el Uso de los Resultados de Evaluaciones Educativas



Informe final



Estudios e  
investigaciones

**INEE**  
Instituto Nacional para la  
Evaluación de la Educación  
México

**Experiencias Internacionales sobre el  
Uso de los Resultados de Evaluaciones Educativas**  
Primera edición, 2016

**Coordinador**

Andrés Sandoval-Hernández (México)

**Autores**

Sara Sofía Calvario Ruiz (México)  
Diego Alonso Carrasco Ogaz (Chile)  
Francisco Javier García Crespo (España)  
Daniela Patricia Jiménez Scheuch (Chile)  
María Guadalupe Pérez Martínez (México)  
David Esteban Torres Iribarra (Chile)

D. R. © Instituto Nacional para la Evaluación de la Educación  
Barranca del Muerto 341, San José Insurgentes, Benito Juárez, 03900, Ciudad de México.

**Corrección de estilo**

Carlos Garduño González  
Hugo Soto de la Vega

**Diseño**

Heidi Puon Sánchez

**Formación**

Martha Alfaro Aguilar

Hecho en México  
Distribución gratuita. Prohibida su venta.  
Consulte el catálogo de publicaciones en línea: [www.inee.edu.mx](http://www.inee.edu.mx)

Publicación a cargo de la Dirección General de Investigación e Innovación. 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 por cualquier sistema mecánico o electrónico para fines no comerciales. Cítese de la siguiente manera:

Sandoval-Hernández, Andrés (Coord.) (2016). *Experiencias Internacionales sobre el Uso de los Resultados de Evaluaciones Educativas*. México: INEE.

## DIRECTORIO

### JUNTA DE GOBIERNO

Sylvia Irene Schmelkes del Valle  
CONSEJERA PRESIDENTA

Eduardo Backhoff Escudero  
CONSEJERO

Gilberto Ramón Guevara Niebla  
CONSEJERO

Margarita María Zorrilla Fierro  
CONSEJERA

Teresa Bracho González  
CONSEJERA

### TITULARES DE UNIDAD

Francisco Miranda López  
UNIDAD DE NORMATIVIDAD Y POLÍTICA EDUCATIVA

Jorge Antonio Hernández Uralde  
UNIDAD DE EVALUACIÓN DEL SISTEMA EDUCATIVO NACIONAL

Carmen Reyes Guerrero  
UNIDAD DE INFORMACIÓN Y FOMENTO DE LA CULTURA  
DE LA EVALUACIÓN

Susana Justo Garza (encargada)  
UNIDAD DE PLANEACIÓN, COORDINACIÓN  
Y COMUNICACIÓN SOCIAL

Miguel Ángel de Jesús López Reyes  
UNIDAD DE ADMINISTRACIÓN

Luis Felipe Michel Díaz  
CONTRALOR INTERNO

José Roberto Cubas Carlín  
COORDINACIÓN DE DIRECCIONES DEL INEE  
EN LAS ENTIDADES FEDERATIVAS

**Dirección General de Difusión  
y Fomento de la Cultura de la Evaluación**  
José Luis Gutiérrez Espíndola

**Dirección de Difusión y Publicaciones**  
Alejandra Delgado Santoveña

## Agradecimientos

---

Este proyecto de investigación no hubiera sido posible sin la iniciativa y el acompañamiento de Alberto Navarrete Zumárraga, Director General de Investigación e Innovación del INEE. Los autores desean agradecer a Martha Chicharro Gutiérrez, Subdirectora de Vinculación para la Investigación e Innovación del mismo instituto, por la lectura crítica de versiones preliminares de este manuscrito y por sus valiosos comentarios para mejorarlo. Para la realización de este trabajo contamos también con la lectura y el apoyo logístico y administrativo de Enrique Lira Hernández y Omar Cervantes Olivares. Finalmente agradecemos a Teresa Bracho González, Consejera de la Junta de Gobierno del INEE, por la lectura de este documento y por la elaboración de la presentación del mismo, que sin duda sitúa este trabajo en perspectiva y ofrece elementos de contexto que contribuyen a ampliar su marco de interpretación.

# Índice

8	<b>Presentación</b>
11	<b>Introducción</b>
12	<b>Plan Nacional para la Evaluación de los Aprendizajes (PLANEA)</b>
14	<b>Metodología</b>
15	Identificación de prácticas internacionales
17	Caracterización de las prácticas internacionales seleccionadas
18	Síntesis y conclusiones
19	<b>1. El caso de Alemania</b>
	1.1. Características de la evaluación de gran escala
	1.1.1. Introducción
20	1.1.2. Marco administrativo/legal
	1.1.3. Procedimientos y metodología
21	1.2. Características de las prácticas de uso de los resultados de las evaluaciones
	1.2.1. Introducción
23	1.2.2. Herramientas para apoyar el uso de las evaluaciones
25	1.2.3. Uso de los resultados
26	1.2.4. Opinión pública
	1.2.5. Ejemplos de materiales de apoyo
27	<b>2. El caso de Chile</b>
	2.1. Características de la evaluación de gran escala
	2.1.1. Introducción
28	2.1.2. Marco administrativo/legal
29	2.1.3. Procedimientos y metodología
30	2.2. Características de las prácticas de uso de los resultados de las evaluaciones
	2.2.1. Introducción
33	2.2.2. Herramientas para apoyar el uso de las evaluaciones
37	2.2.3. Uso de los resultados
38	2.2.4. Opinión pública
40	2.2.5. Ejemplos de materiales de apoyo
41	<b>3. El caso de España</b>
	3.1. Características de la evaluación de gran escala
	3.1.1. Introducción

42	3.1.2. Marco administrativo/legal
43	3.1.3. Procedimientos y metodología
44	3.2. Características de las prácticas de uso de los resultados de las evaluaciones
	3.2.1. Introducción
45	3.2.2. Herramientas para apoyar el uso de las evaluaciones
50	3.2.3. Uso de los resultados
	3.2.4. Opinión pública
51	3.2.5. Ejemplos de materiales de apoyo
53	<b>4. El caso de Estados Unidos de América</b>
	4.1. Características de la evaluación de gran escala
	4.1.1. Introducción
54	4.1.2. Marco administrativo/legal
55	4.1.3. Procedimientos y metodología
56	4.2. Características de las prácticas de uso de los resultados de las evaluaciones
	4.2.1. Introducción
58	4.2.2. Herramientas para apoyar el uso de las evaluaciones
	Explorador de datos NAEP ( <i>NAEP Data Explorer, NDE</i> )
	Ejemplos de preguntas ( <i>sample questions</i> )
60	Comparaciones estatales ( <i>State comparisons</i> )
	Perfiles estatales ( <i>State profiles</i> )
	Perfiles de los distritos ( <i>District profiles</i> )
61	Bases de datos para análisis secundarios
	Recursos en otras páginas web ( <i>The Nation's Report Card</i> )
	4.2.3. Uso de los resultados
62	4.2.4. Opinión pública
63	4.2.5. Ejemplos de materiales de apoyo
64	<b>5. El caso de PISA for Schools</b>
	5.1. Características de la evaluación a gran escala
	5.1.1. Introducción
65	5.1.2. Marco administrativo/legal
66	5.1.3. Procedimientos y metodología
68	5.2. Características de las prácticas de uso de los resultados de las evaluaciones
	5.2.1. Introducción
69	5.2.2. Herramientas para apoyar el uso de las evaluaciones
71	5.2.3. Uso de los resultados
72	5.2.4. Opinión pública
75	5.2.5. Ejemplos de materiales de apoyo
79	<b>Síntesis y conclusiones</b>
85	<b>Referencias</b>

# Índice de figuras

## Figuras

- 23 Figura 1. Modelo de uso de resultados sugerido por la agencia de calidad escolar de Nordrhein-Westfalen (QUA-LIS NRW)
- 31 Figura 2. Calendario de aplicaciones por curso y área curricular
- 33 Figura 3. Extracto de la página [www.simce.cl](http://www.simce.cl)
- 34 Figura 4. Diagrama de indicadores que son presentados a la prensa
- 36 Figura 5. Ejemplo de las recomendaciones respecto a cómo usar y como no usar los resultados reportados
- 39 Figura 6 Artículo de un periódico nacional publicado con *ranking* de colegios acorde a puntajes de las pruebas SIMCE de segundo básico
- 76 Figura 7 Comparación de los resultados de la escuela, a nivel local e internacional
- 77 Figura 8 Perfiles de lectores en PISA for Schools
- 78 Figura 9 Porcentaje de lectores en la escuela y a nivel nacional (avisos, mensajes, llamadas, entre otros)

# Presentación

Las evaluaciones de logro en el terreno educativo han sido concebidas desde su origen como instrumentos que den cuenta de los aprendizajes logrados por los evaluados. Con ello se obtiene una dimensión, en el nivel individual, de lo alcanzado por un estudiante y, en el social, de los resultados de los sistemas educativos en su conjunto. Con frecuencia se encuentra en la literatura especializada, en los medios y en los discursos de políticos y funcionarios de la educación, que las evaluaciones de logro son necesarias para saber qué sucede y para orientar el curso de las acciones a futuro, ya sea para consolidar logros o para corregir rumbos.

Producto de nuestra preocupación por los usos de la evaluación educativa a lo largo de los años, tanto desde la trinchera de la investigación como de la evaluación de políticas educativas, hemos desarrollado una convicción: para que las promesas de esos instrumentos se traduzcan en consecuencias afortunadas se requiere mucho más que pulcritud metodológica. Para que sirva de algo el tiempo, el esfuerzo y los recursos invertidos en su realización, es necesario que la información que ofrecen sea usada efectiva y adecuadamente por quienes puedan servirse de su contenido. La referencia al uso de los resultados de las evaluaciones hace hincapié en “el conjunto de acciones deliberadas, ejecutadas por diferentes receptores o usuarios, que retoman (en diferentes grados y para distintos propósitos) los resultados, los métodos y los conceptos derivados de la evaluación educativa, como uno de los insumos relevantes para sustentar sus decisiones y actividades en su contexto particular de actuación” (Bracho, 2016).<sup>1</sup> Para analizar los usos específicos, es imprescindible considerar el área de intervención particular del asunto, el objeto y el contexto de la evaluación.

En la literatura especializada abundan referencias a la forma de construir, e incluso de interpretar correctamente, estos instrumentos y sus resultados. Pero investigaciones que atiendan de manera puntual el problema de los usos de estos resultados no son tan abundantes, a pesar de que el tema ha cobrado relevancia capital en los últimos años. Es por ello que la investigación que aquí se presenta resulta particularmente valiosa, tanto en el contexto general de la educación mexicana como en la dimensión específica de la labor del Instituto Nacional para la Evaluación de la Educación, en especial, aunque no exclusivamente, en lo que atañe al Plan Nacional para la Evaluación de los Aprendizajes, PLANEA.

<sup>1</sup> Bracho, T. (2016) Usos efectivos para la mejora: ¿cómo hacemos para que esto ocurra?, en *Gaceta de la Política Nacional de Evaluación Educativa en México*, México, Instituto Nacional para la Evaluación de la Educación, núm. 4, marzo-junio 2016, pp.38-42

Los usos de las evaluaciones de logro forman parte ya del mandato legal conferido al INEE, y no son una mera preocupación intelectual o académica. De ahí la importancia de reunir en un solo volumen una revisión sistemática de las experiencias de distintos países (Alemania, Chile, España, Estados Unidos) y del caso de *PISA para escuelas*, y poner al alcance del lector información sobre las características generales de las evaluaciones —marcos legales y administrativos, objetivos, diseños, metodologías de levantamiento de los datos, entre otros—, los usos pretendidos de los resultados, así como el análisis tanto de los usuarios de los resultados como de los propósitos mismos de las evaluaciones. Colocar esta información ante los ojos de funcionarios y público interesado en el tema, sin duda enriquecerá el repertorio de prácticas a considerar para convertir al Plan Nacional para la Evaluación de los Aprendizajes (PLANEA) en un instrumento efectivo de mejora en el ámbito de las escuelas mexicanas.

De las buenas prácticas analizadas sobresalen iniciativas de las que puede aprenderse mucho, entre ellas:

- La elaboración de informes en dos tiempos: preliminar y definitivo. Con esto se logra superar una de las críticas más frecuentes a la evaluación, la relativa a la oportunidad con la que se ponen a disposición del público los resultados.
- El desarrollo de materiales de apoyo para interpretar los informes, con la intención de favorecer el desarrollo de habilidades en los públicos usuarios, así como advertir sobre los usos inadecuados de los resultados.
- Las visitas de personal capacitado a las escuelas para explicar y mostrar cómo realizar la interpretación de los informes para su utilización efectiva en procesos de mejora escolar.
- La liberación de los ítems de las pruebas para que sean utilizados como material didáctico en las escuelas.
- La promoción explícita de los marcos de referencia de las evaluaciones para el diseño, evaluación y revisión del currículo.
- La promoción de la discusión de los informes en las escuelas evaluadas, para facilitar su comprensión en la mejora de los aprendizajes.
- La creación de sitios web con recursos para apoyar la interpretación de las pruebas.
- El uso de redes sociales para estimular el intercambio entre las diversas audiencias, con la intención de promover y facilitar la crítica constructiva a las acciones realizadas, relacionadas tanto con las evaluaciones de logro como con los usos de sus resultados.

En México aún nos falta camino por recorrer para llegar al punto en donde nuestras evaluaciones de logro atiendan todos los aspectos relevantes y se ocupen de todos los usuarios posibles, de manera tal que conduzcan a verificar las mejoras en la calidad y distribución de la educación. Sin embargo, acelerar hoy el paso y concentrar esfuerzos en este objetivo final, razón de ser de nuestras evaluaciones, es un punto crítico relevante por las siguientes razones:

**Oportunidad.** Nunca como ahora ha sido tan urgente generar un vínculo claro entre la evaluación y la mejora del Sistema Educativo Mexicano. Que la evaluación sea el mecanismo para dar cuenta del derecho a una educación de calidad, como fue establecido en la reforma constitucional de 2013, requiere que nos preocupemos tanto por la independencia y el rigor metodológico de las evaluaciones, cuanto por los usos efectivos de la información que producen.

**Originalidad.** Contar con un estudio que reúna una descripción de prácticas tan significativas como las que se presentan permite aprender de lo realizado, lo cual es a la vez una necesidad y un espacio de desarrollo.

**Relevancia.** Muchos de los mejores ejemplos de buenas prácticas aquí contenidos están ausentes en nuestro contexto, por lo que su identificación y exploración sin duda constituyen un motivo de reflexión e inspiración.

Por ello, hacemos una entusiasta invitación a la lectura de *Experiencias Internacionales sobre el uso de los Resultados de Evaluaciones Educativas*, mismo que auguramos resultará de mucho interés para los diversos públicos interesados en esta materia y que esperamos sea de provecho para la construcción de una nueva y sólida experiencia de fomento de usos efectivos de las evaluaciones de logro en México para la mejora educativa.

**Dra. Teresa Bracho González**  
**Consejera de la Junta de Gobierno**

# Introducción

Este documento describe una muestra de prácticas internacionales de uso de los resultados de las evaluaciones de logro educativo para dar realimentación a las escuelas. El objetivo general es contribuir al desarrollo de un modelo conceptual de uso del Plan Nacional para las Evaluaciones del Aprendizaje (PLANEA) para las escuelas mexicanas, y su cumplimiento se guía por tres objetivos específicos:

- Realizar una revisión de prácticas internacionales en materia de uso de los resultados de las evaluaciones de logro educativo con el fin de seleccionar una muestra que pueda proporcionar información útil para el uso de los resultados de PLANEA.
- Desarrollar un documento que describa detalladamente al menos cinco prácticas internacionales en materia de uso de evaluaciones educativas de gran escala para dar realimentación a las escuelas que participan en ellas.
- Analizar la descripción de las prácticas internacionales a la luz del diseño de PLANEA con el fin de identificar un conjunto de opciones de prácticas y políticas que se juzguen potencialmente aplicables en el contexto del sistema educativo mexicano.

Para ello, además de esta introducción, el presente informe contiene ocho secciones principales. En la primera se incluye una breve descripción de PLANEA, sus antecedentes, las áreas que evalúa y sus objetivos generales.

La segunda sección corresponde a la exposición de la metodología utilizada en este documento. En ella se describe el procedimiento para seleccionar las prácticas internacionales en materia de uso de evaluaciones educativas de gran escala, a fin de dar realimentación a las escuelas incluidas en este reporte; asimismo, se presentan las características que se consideraron de dichas prácticas para realizar el análisis que permitió identificar un conjunto de acciones y políticas potencialmente aplicables para el uso de la información generada por PLANEA.

Las secciones que van de la tercera a la séptima contienen el análisis sistemático de cinco prácticas internacionales: Alemania, Chile, España, Estados Unidos de América y PISA para Escuelas. Cada una de estas secciones se divide a su vez en dos partes. En la primera, se describen las características generales de las evaluaciones, por ejemplo, los marcos legales y administrativos, objetivos, diseños, metodologías de levantamiento de datos, etcétera. En la segunda, se exponen los usos pretendidos de las evaluaciones y las herramientas disponibles para ello; se presenta también una compilación de consideraciones sobre los usos de resultados de las evaluaciones en la opinión pública, y algunos ejemplos de materiales de apoyo disponibles. Finalmente, la última sección se dedica a las conclusiones y recomendaciones.

# Plan Nacional para la Evaluación de los Aprendizajes (PLANEA)

De acuerdo con el documento rector de PLANEA (INEE, 2015a), con el fin de conocer qué tanto los estudiantes mexicanos dominan aprendizajes clave en diferentes momentos de la educación obligatoria, el Instituto Nacional para la Evaluación de la Educación (INEE) desarrolló PLANEA con miras a obtener información complementaria para apoyar la mejora educativa. En su elaboración se ha considerado como antecedente el análisis de las fortalezas y limitaciones de los Exámenes de la Calidad y el Logro Educativos (EXCALE) y de la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE), y se atienden las recomendaciones derivadas de un estudio que, durante 2013 y 2014 y por petición del INEE, llevó a cabo un grupo de especialistas de diferentes instituciones (Martínez Rizo *et al.*, 2015).

De este modo, se proyecta que los resultados de las evaluaciones de PLANEA sirvan para apoyar la mejora educativa a partir de:

- Informar a la sociedad sobre el estado que guarda la educación en términos del logro de aprendizaje de los estudiantes y de la equidad (o inequidad) que existe en los resultados educativos.
- Aportar a las autoridades educativas información relevante para el monitoreo, la planeación, la programación y la operación del sistema educativo y sus centros escolares.
- Ofrecer información pertinente, oportuna y contextualizada a las escuelas y a los docentes que ayude a mejorar sus prácticas de enseñanza y el aprendizaje de todos sus estudiantes.
- Contribuir al desarrollo de directrices para la mejora educativa con información relevante sobre los resultados educativos y los contextos en que se dan.

Es importante mencionar que el documento rector de PLANEA estipula que los resultados de las evaluaciones no deberán utilizarse para juzgar el desempeño de los docentes, realizar *rankings* de escuelas ni justificar procesos punitivos u otros de control administrativo sobre estudiantes, docentes o escuelas.

PLANEA se puso en operación a partir del ciclo escolar 2014-2015 en coordinación con la Secretaría de Educación Pública (SEP). Su diseño considera la evaluación de competencias de lenguaje y comunicación, matemáticas y habilidades socioemocionales en tres modalidades:

- Evaluación del Logro del Sistema Educativo Nacional (ELSEN), aplicada por el INEE. Esta prueba permite evaluar una cantidad amplia de aprendizajes clave y se aplica a muestras representativas de estudiantes de todo el país que cursan los grados terminales de la educación obligatoria (tercero de preescolar, sexto de primaria, tercero de secundaria y el último grado de bachillerato).
- Evaluación de Logro en Centros Escolares (ELCE), aplicada por la SEP. Evalúa una cantidad reducida de contenidos académicos, y se aplica anualmente en todas las escuelas del país en los grados terminales de primaria, secundaria y educación media superior.
- Evaluación Diagnóstica Censal (EDC) (aplicada por la escuela y los docentes). Para su aplicación se entregaron pruebas estandarizadas a los docentes de cuarto grado de primaria, al inicio del curso 2015-2016. Se espera que, al finalizar la aplicación, la comunidad escolar (el consejo técnico escolar y el consejo escolar de participación social) analice los resultados a partir de una guía, y los utilice como elementos para planear el trabajo en el aula.

En junio de 2015 el INEE y la SEP aplicaron de manera concurrente la ELSEN y la ELCE, respectivamente. La muestra de la primera constó de 104 204 estudiantes de sexto de primaria distribuidos en 3 446 escuelas y 144 517 estudiantes de tercero de secundaria pertenecientes a 3 529 centros escolares (INEE, 2015b). Por su parte, la muestra de la ELCE constó de 1 638 158 estudiantes de sexto de primaria distribuidos en 73 769 escuelas, y 1 007 415 estudiantes pertenecientes a 31 635 escuelas (SEP, 2015).

Durante los primeros meses del ciclo escolar 2015-2016 cada una de las escuelas que participó en la aplicación de PLANEA recibió un reporte de contextualizado de la ELCE (ver <http://planea.sep.gob.mx/>). De manera simultánea, el INEE publicó los resultados de la ELSEN. Se tiene previsto que estos resultados permitan realizar un diagnóstico del Sistema Educativo Nacional, en el caso de la ELSEN, así como alentar la reflexión de las comunidades educativas para contribuir a la mejora de los procesos de enseñanza aprendizaje en las escuelas, en el caso de la ELCE.

# Metodología

En esta sección se presentan los procedimientos seguidos con el fin de alcanzar los objetivos mencionados anteriormente. En términos generales, estos pasos consistieron en identificar prácticas internacionales en materia de uso de evaluaciones educativas de gran escala para realimentar a las escuelas que además pudieran servir como casos de estudio para este proyecto en caracterizar las prácticas internacionales seleccionadas, y en analizar dichas caracterizaciones con el propósito de identificar un conjunto de opciones de prácticas y políticas que se juzguen potencialmente aplicables para el uso de la información generada por el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). Sin embargo, antes de describir con detalle estos pasos, es necesario delimitar nuestro objeto de interés con mayor precisión.

La evaluación educativa ciertamente no es un fenómeno nuevo. Está ligada a la aparición de la figura del maestro, quien siempre ha tenido la evaluación de los avances de cada alumno como una de sus funciones primordiales. Lo que sí es un fenómeno reciente, como apunta Martínez Rizo (2009), es la existencia de evaluaciones a gran escala. Para este proyecto, estas evaluaciones se definen como aquellas en las que se aplican pruebas estandarizadas a grandes números de alumnos, con el fin de valorar el nivel de aprendizaje que alcanzan los estudiantes en el sistema educativo de un país, una región, un distrito o una modalidad educativa (Martínez Rizo, 2009).

Ahora bien, en cuanto a los usos que se le han dado a la información que se obtiene de este tipo de evaluaciones, autores como McLaughlan de Arregui (2012) sugieren los siguientes:

- Elevar la conciencia ciudadana sobre los retos de la educación.
- Mejorar el diseño de las políticas educativas y la asignación de recursos.
- Diseñar programas de apoyo y seleccionar escuelas o grupos de estudiantes para participar en ellos.
- Realizar investigaciones y evaluar el impacto de los programas.
- Revisar expectativas curriculares o establecer estándares de aprendizaje.
- Promover que se asuma responsabilidad por los resultados, la rendición de cuentas y la libre elección de escuelas.
- Mejorar la gestión pedagógica e institucional en las escuelas o áreas escolares.

Aunque los diferentes usos que se mencionan son todos importantes, el análisis que se desarrolla en este trabajo se relaciona más con el último punto, es decir, con el uso de las evaluaciones a gran escala para dar realimentación a las escuelas.

## Identificación de prácticas internacionales

Para identificar las prácticas internacionales que se analizaron, fue necesario identificar un conjunto de conceptos básicos para guiar su selección y su caracterización. La descripción y la definición de las características clave consideradas en este proyecto constituyen este conjunto de conceptos.

Se hizo una exploración de los países y organismos internacionales que podrían ser considerados en el proyecto. Inicialmente se buscó tener diversidad geográfica, de las características de las evaluaciones a gran escala de que se tratase, y de las de los usos pretendidos de estas evaluaciones. Las características clave que se consideraron para la selección de casos se definieron utilizando como referencia otros estudios similares (ver, por ejemplo, OECD, 2013), y son las siguientes.

En cuanto a las características de las evaluaciones a gran escala:

- **Marco administrativo/legal.** Se refiere a identificar qué organismo está a cargo de la aplicación de la evaluación de que se trate, por ejemplo, el ministerio de educación nacional, una agencia gubernamental subnacional (estatal, regional, distrital), una agencia descentralizada/independiente, un organismo internacional. Información adicional dentro de este apartado se refiere a la descripción del marco legal de la evaluación, es decir, si está regulada por la legislación del país, si es de participación voluntaria, etcétera.
- **Procedimientos y metodología.** La metodología y los procedimientos que se utilizan para el desarrollo y la aplicación de una evaluación son también elementos fundamentales para su caracterización. En este apartado se considerarán, por ejemplo:
  - Los actores involucrados en el desarrollo y la aplicación de la evaluación (como padres, alumnos, maestros, sindicatos de maestros, académicos, personal del ministerio de educación u otros organismos gubernamentales, hacedores de política pública, organismos internacionales, etcétera).
  - El tipo de diseño muestral utilizado (como estratificado, por etapas, etcétera), o bien si el estudio es censal.
  - La metodología para el levantamiento de datos (quién/cómo aplica las pruebas, quién/cómo se generan las bases de datos correspondientes, etcétera).
  - La población objetivo de la evaluación (como grados, edades, etcétera).
  - El objeto de la evaluación (asignaturas, temas, contenidos, etcétera).
- **Disponibilidad de información** sobre la evaluación, su desarrollo, su aplicación, sus resultados, los manuales para el uso de los datos que genera, etcétera.

En cuanto a las características de los usos pretendidos de los resultados:

- **Usuarios de los resultados.** Quiénes, de acuerdo con la documentación relativa a la evaluación, son los actores que se pretende que usen la información que se genera (como, alumnos, maestros, padres de familia, directores de escuela, supervisores, hacedores de política, etcétera).

- **Propósito de la evaluación.** Se refiere a los usos planeados para los resultados de la evaluación (como acreditación o certificación, ordenamiento y selección, toma de decisiones, incentivos y sanciones, rendición de cuentas, etcétera).

Como ya se mencionó, uno de los criterios considerados para la selección de los casos incluidos en este proyecto fue la diversidad geográfica, de modo que se hizo un análisis preliminar a partir del siguiente listado de países y organismos internacionales:

- América del Norte: Canadá, Estados Unidos de América.
- Latinoamérica: Chile, Perú.
- Europa: España, Alemania.
- África: Botsuana, Sudáfrica.
- Asia: Japón, Singapur.
- Oceanía: Australia, Nueva Zelanda.
- Organismos internacionales: PISA for Schools (OCDE),<sup>1</sup> School level Reports sobre TIMSS y PIRLS (IEA).<sup>2</sup>

Se revisaron fuentes electrónicas disponibles en los idiomas inglés, español y alemán sobre los sistemas de evaluación de cada uno de los países u organismos internacionales para valorar la suficiencia de información sobre sus evaluaciones a gran escala y sobre los usos de la información que éstas generan para dar realimentación a las escuelas.

Las fuentes electrónicas consultadas fueron: páginas web de los sistemas educativos o las agencias de evaluación de la calidad de la educación en los países y organismos internacionales mencionados; así como artículos de investigación y artículos de prensa.

A partir de este análisis preliminar, tomando en cuenta la disponibilidad de la información pública para caracterizar las diferentes prácticas de acuerdo con el marco conceptual, se eligieron los siguientes casos:

- Estados Unidos de América.
- Chile.
- España.
- Alemania.
- PISA for Schools.

Es importante señalar que de ninguna manera se sugiere que las prácticas incluidas en este documento sean modelos a seguir en materia de uso de evaluaciones a gran escala para dar realimentación a escuelas, al menos no sin haber analizado cuidadosamente el contexto en el que han sido empleadas y la (di)similitud entre éste y el del sistema educativo mexicano. Estos casos meramente ilustran algunos ejemplos de mecanismos y procedimientos que han sido empleados

<sup>1</sup> PISA for Schools es una herramienta de evaluación para escuelas basada en el Programme for International Student Assessment (PISA) que organiza la Organización para la Cooperación y el Desarrollo Económicos (OCDE).

<sup>2</sup> Los School Level Reports sobre TIMSS and PIRLS son reportes para las escuelas participantes en los estudios Trends in International Mathematics and Science Study (TIMSS) y Progress in International Reading Literacy Study (PIRLS), organizados por la International Association for the Evaluation of Educational Achievement (IEA).

en otros lugares para hacer uso de la información generada mediante evaluaciones a gran escala para realimentar a las escuelas.

## Caracterización de las prácticas internacionales seleccionadas

Como el nombre de este apartado lo sugiere, una vez identificadas las prácticas internacionales, el siguiente paso consistió en caracterizarlas. La información utilizada para ello se recabó utilizando un formato para registro de prácticas.

Existe un cuerpo de investigación considerable sobre el uso de la información generada por procesos evaluativos. En este sentido, la literatura especializada sugiere que la relación entre la información producida por las evaluaciones a gran escala y su uso por los diferentes actores no es simple (Sandoval-Hernández y Neuschmidt, 2014). De hecho, existe evidencia de que los usos de este tipo de información son más bien múltiples y dependen en gran medida de diversos factores, como el nivel de complejidad de la información, el contexto político, las características personales de quien recibe la información, la disponibilidad de apoyos para su uso, etcétera (ver, por ejemplo, Elmore, 2008; Hein y Taut, 2010). Por ello, una vez seleccionadas las prácticas internacionales, se procedió a caracterizarlas de acuerdo con los criterios que se describen abajo. Estas categorías analíticas buscan, en la medida de lo posible, capturar la complejidad de los factores que dificultan o promueven el uso de la información generada por las evaluaciones.

- **Descripción.** Este paso comienza por proporcionar una descripción detallada tanto de “las características de la evaluación a gran escala” como del “uso que se hace de sus resultados”, con especial hincapié en el uso de éstos para realimentar a las escuelas. El contenido de dicha descripción reflejará los criterios utilizados para seleccionar las prácticas internacionales que se describen en el paso anterior.
- **Herramientas para apoyar el uso de las evaluaciones.** Esta sección incluye una descripción de los apoyos existentes para hacer uso de los resultados de las evaluaciones a gran escala, por ejemplo, bases de datos, guías o manuales de usuario, sitios web, herramientas gráficas, reportes escolares, organización de eventos, talleres, cursos, folletos, etcétera. Para cada herramienta identificada se presenta información sobre los actores involucrados en su diseño, los usuarios pretendidos, los objetivos particulares de la herramienta, las estrategias utilizadas para promover su uso, la accesibilidad y el nivel de complejidad para su uso, etcétera.
- **Uso de los resultados.** Esta sección recoge evidencia del uso y (cuando sea posible) de los resultados del empleo de los recursos descritos anteriormente. Es decir, se refiere, por ejemplo:
  - A la recopilación de estadísticas de uso de las herramientas de apoyo identificadas.
  - A la recopilación y el análisis de estudios evaluativos de las prácticas de uso de las evaluaciones a gran escala para realimentar a las escuelas.

Los tipos de documentos que se incluyen son, por un lado, estadísticas y reportes generados por las agencias evaluadoras u otros organismos independientes, y, por otro lado, artículos publicados en revistas académicas o de divulgación en círculos especializados, y reportes que hagan evaluaciones independientes de dichas prácticas.

- **Opinión pública.** Esta sección contiene consideraciones sobre los usos de resultados de las evaluaciones en la opinión pública. Es decir, información publicada en medios masivos de información (como prensa, radio —*podcasts*—, televisión —*videos*—, etcétera), en redes sociales en línea (como Facebook, Twitter, LinkedIn, etcétera), en blogs especializados, en los comentarios a las versiones en línea de notas periodísticas, entre otros. La búsqueda de información se centró en fortalezas y debilidades identificadas por la opinión pública, y en las críticas y los elogios hechos para diferentes herramientas o prácticas de uso de los resultados de las evaluaciones.
- **Ejemplos de materiales de apoyo.** Esta sección incluye ejemplos de los materiales empleados para la elaboración de la caracterización de prácticas. Por ejemplo (cuando los documentos estaban disponibles), se incluyen muestras de los reportes distribuidos a las escuelas y de los folletos u otros materiales promocionales, resúmenes de ponencias, actas de conferencias y seminarios, temarios de seminarios y talleres, impresiones en papel de herramientas web (*screenshots*), así como una lista de direcciones de páginas de Internet, videos y cuentas de redes sociales relevantes.

## Síntesis y conclusiones

La última etapa del proyecto consiste en analizar los hallazgos a la luz del diseño de PLANEA. La intención fue utilizar la información presentada en las secciones anteriores para identificar un conjunto de opciones de prácticas y políticas potencialmente aplicables para el uso de la información generada por PLANEA para realimentar a las escuelas participantes en el contexto del sistema educativo mexicano.

# 1

## El caso de Alemania

---

### 1.1. Características de la evaluación de gran escala

#### 1.1.1. Introducción

Las pruebas nacionales alemanas VERA (acrónimo de *Vergleichsarbeiten*, que en algunos estados han adoptado otros nombres) comenzaron a aplicarse durante el año escolar 2007/2008 en algunos estados federales, en el marco de la línea de acción “desarrollo de pruebas comparativas entre estados para la revisión de la efectividad escolar”, como parte de la “estrategia global de monitoreo del sistema escolar alemán” propuesta en junio de 2006 por el Consejo de Ministerios de Educación y Asuntos Culturales (KMK, por sus siglas en alemán) de los 16 estados que componen la República Federal Alemana. En las actualizaciones efectuadas a esta “estrategia global” en 2015, se enfatizó el papel de VERA en el contexto de los “procedimientos de aseguramiento de la calidad a nivel escuela”, entendiendo que las pruebas forman parte de una serie de medidas que permiten a los estados federales implementar acciones de desarrollo y mejoramiento escolar basadas en evidencia.<sup>1</sup> Se aplican en tercer y octavo grados, y sus tres principales objetivos son los siguientes:

- **Desarrollo escolar y de las prácticas de aula:** las pruebas pretenden potenciar aspectos disciplinarios, didácticos y pedagógicos a nivel escuela. La participación activa de los profesores en la aplicación y la corrección de las pruebas pretende promover cooperación y discusión acerca de los estándares de enseñanza, planificación de prácticas de aula y prácticas de evaluación.
- **Evaluación del cumplimiento de los estándares de enseñanza:** es importante destacar que estos estándares en Alemania son relativamente recientes. Con el fin de poner a disposición de todos los estados un marco de referencia para la calidad de la enseñanza escolar, los estados federales establecieron estándares nacionales de enseñanza para el último año de cada ciclo escolar, que fueron aprobados en 2003, 2004 y 2012 para los grados cuarto, noveno y décimo, y para el final de la enseñanza secundaria), y han sido adoptados gradualmente, por lo que las pruebas comparativas son un instrumento esencial para evaluar su implementación.

<sup>1</sup> Extraído de *The Education System in the Federal Republic of Germany 2012/2013. A Description of the Responsibilities, Structures and Developments in Education Policy for the Exchange of Information in Europe*. Recuperado de: [http://www.kmk.org/fileadmin/doc/Dokumentation/Bildungswesen\\_en\\_pdfs/dossier\\_en\\_ebook.pdf](http://www.kmk.org/fileadmin/doc/Dokumentation/Bildungswesen_en_pdfs/dossier_en_ebook.pdf)

- **Medición y mejoramiento diagnóstico** (sólo para VERA 3):<sup>2</sup> antes de la aplicación de las pruebas, los profesores realizan una estimación de la proporción de respuestas correctas que se obtendrán en la clase en que se aplicará la prueba, que posteriormente se contrasta con los resultados obtenidos.

Si bien los objetivos aquí señalados apuntan directamente al uso de resultados por parte del profesor, se espera que otros actores del sistema, tales como integrantes del equipo directivo e inspectores escolares, también los usen, enfatizando que su propósito es el desarrollo y el mejoramiento escolar.

### 1.1.2. Marco administrativo/legal

Las pruebas comparativas alemanas se desarrollan en conjunto entre el Instituto para el Mejoramiento de la Calidad Educativa (IQB, por sus siglas en alemán) y las unidades ministeriales o agencias de calidad encargadas en cada uno de los estados federales. El IQB es un instituto académico alojado en la Universidad Humboldt en Berlín, el cual fue fundado en 2004 por el Consejo de Ministerios de Educación para apoyar a los 16 estados federales en el mejoramiento y el aseguramiento de la calidad del sistema escolar.

El Centro de Investigación Empírica en Educación (ZEPF, por sus siglas en alemán), de la Universidad Koblenz-Landau, también participa en VERA. En un comienzo tuvo un papel relevante en la construcción de las pruebas de aprendizaje, y en la actualidad ofrece la máscara de digitación de las pruebas a algunos estados, así como la descarga de documentos asociados a la entrega de resultados (ver ZEPF, 2014). Sin embargo, no aparece mencionado explícitamente en los documentos oficiales del KMK como participante.

La participación en VERA está regulada por decretos legales de cada estado federal, y es obligatoria para todos los estudiantes de los grados evaluados, aunque existen excepciones para aquellos estudiantes que poseen alguna necesidad educativa especial, requieren algún apoyo especial para llevarla a cabo o han vivido menos de 12 meses en Alemania y no dominan el idioma. La participación de las escuelas privadas reconocidas por el estado es voluntaria.

### 1.1.3. Procedimientos y metodología

El IQB es responsable de la construcción de los ítems que componen las pruebas, los cuales son elaborados por profesores y revisados por expertos en la disciplina y la didáctica, así como por especialistas en medición, luego de lo cual son sometidos a pilotaje. La selección de ítems y el establecimiento de su nivel de dificultad, así como el ensamblaje de las pruebas definitivas y la elaboración de material didáctico complementario también se efectúan en el IQB. Cabe señalar que existe la posibilidad de realizar adecuaciones a la extensión de los cuadernillos de prueba y la inclusión de ciertos ítems en particular, de acuerdo con solicitudes específicas realizadas por los estados.

---

<sup>2</sup> Extraído de <http://vera-web.uni-landau.de/verapub/>

Cada estado federal, por su parte, es responsable de todo lo relativo a la aplicación, es decir, la impresión y la distribución de los cuadernillos de prueba, la corrección y el ingreso de datos, los análisis estadísticos, el diseño de los documentos de resultados e instancias de retroalimentación, la entrega de resultados, y el apoyo a los colegios con acciones complementarias asociadas a ésta. Cabe destacar que la aplicación en las escuelas es hecha por los propios maestros, y que ellos son también quienes corrigen las pruebas e ingresan los datos.

El estudio es censal, y en él participan todos los estudiantes de tercer y octavo grados de los 16 estados federales alemanes.

Las asignaturas evaluadas son las siguientes:

- En tercer grado alemán y matemáticas.
- En octavo grado alemán, matemáticas, inglés y francés (para estas últimas se ofrecen como primer o segundo idioma extranjero, según sea el caso de cada escuela).

Los estudiantes participantes llevan a cabo obligatoriamente la prueba en una de las asignaturas ofrecidas y, adicionalmente, según lo estipule cada estado federal, puede haber otra(s) asignatura(s) obligatoria(s) o permitirse que la decisión de las asignaturas adicionales que se evalúan quede en manos de las escuelas.

Por ejemplo, en Hessen, para la aplicación 2016 en tercer grado es obligatorio evaluar matemáticas y alemán, mientras que en octavo grado es obligatoria una asignatura opcional, y es voluntario (se recomienda) evaluar una segunda.

Para cada asignatura y ciclo se evalúan las competencias correspondientes según lo determinado en los estándares de enseñanza vigentes, con focos especiales que varían de medición en medición. En el caso de tercer grado, alemán siempre incluye lectura y una competencia adicional, por ejemplo, comprensión auditiva, ortografía o usos del lenguaje. Está prevista la inclusión futura de escritura como opcional. En matemáticas siempre se evalúan dos de las cinco competencias establecidas en los estándares; por ejemplo, en 2016 se evaluarán sólo operaciones y números, y patrones y estructuras.

En el caso de octavo grado en alemán se evalúa siempre lectura y una competencia adicional, que puede ser comprensión auditiva, ortografía, usos del lenguaje o escritura, mientras que en matemáticas se evalúan las cinco competencias que componen los estándares. En idioma extranjero se incluye siempre comprensión lectora y una competencia adicional.

## **1.2. Características de las prácticas de uso de los resultados de las evaluaciones**

### **1.2.1. Introducción**

Desde su creación el propósito de las pruebas comparativas alemanas ha sido orientar el mejoramiento escolar y de las prácticas docentes en las escuelas. Los reportes de resultados que ofrece VERA indican la ubicación de estudiantes en niveles de desempeño, y son los siguientes:

- Reporte de resultados individuales (primera entrega).
- Resultados agregados por curso que incluyen comparación curso-colegio (primera entrega).
- Resultados agregados que incluyen comparación colegio-estado, considerando factores de contexto (segunda entrega).
- Resultados agregados que incluyen comparación de cada curso con un grupo de referencia (segunda entrega).
- Resultados agregados por ítem (segunda entrega).

En 2012 el consejo de ministerios estableció un acuerdo sobre el desarrollo de VERA, e hizo una serie de precisiones para las aplicaciones futuras. A continuación, se detallan aquellas relativas al uso de resultados:

- Se reitera que el principal propósito de VERA, como parte de la estrategia global del KMK, es el mejoramiento escolar y de las prácticas docentes de todas las escuelas alemanas. Complementariamente se señala la función de transmisión del marco disciplinario y didáctico de los estándares de enseñanza. Con el fin de velar por el cumplimiento de ambos propósitos, se acordó que los resultados de VERA no se publicarán, pues esto podría atentar con el propósito de mejoramiento escolar. En caso de presentar resultados en el marco del aseguramiento de la calidad escolar, debe garantizarse que las escuelas se anonimizarán. Este acuerdo impide explícitamente la elaboración de *rankings* de escuelas con los resultados obtenidos en VERA.
- Las pruebas comparativas no consideran la asignación de una nota, no remplazan pruebas sumativas y sus resultados no pueden usarse para pronosticar éxito escolar. Sin embargo, resulta razonable que los propios estudiantes y sus padres obtengan realimentación, la cual debe contextualizarse con las consideraciones pedagógicas que corresponden a la naturaleza de la prueba. Una implicación directa de este acuerdo para los estudiantes de tercer grado es que su desempeño en las pruebas comparativas no puede ser considerado en la recomendación que hace la escuela acerca del tipo de enseñanza secundaria que debe seguir el estudiante.
- Los resultados pueden ser utilizados para propósitos de inspección y supervisión escolar. Cada estado debe decidir cómo se incorporarán indicando reglas claras siempre orientadas al apoyo de los procesos de mejoramiento escolar.
- El documento también incluye recomendaciones respecto del uso de resultados a ser implementadas individualmente por cada estado federal:
  - Fortalecimiento del papel del equipo directivo como responsable de asegurar y mejorar la calidad de la enseñanza realizando un trabajo conjunto con el profesorado basado en los resultados de VERA.
  - Uso de las posibilidades que ofrecen las instituciones locales involucradas en las pruebas para reaccionar con flexibilidad ante las necesidades particulares de cada escuela; por ejemplo, vincular a las escuelas con expertos que las apoyen en la interpretación de los resultados, en la planificación de medidas de apoyo a escuelas o en el diseño de instrumentos de evaluación complementarios a VERA para promover el progreso de los estudiantes individualmente.
  - Poner a disposición de los actores los materiales para maximizar el uso de los resultados, tales como orientaciones para la práctica, generación de una lista de preguntas frecuentes para uso de los datos, y preguntas didácticas que promuevan el intercambio entre profesores.

- Incorporación del uso adecuado de datos de logro de aprendizaje para el desarrollo escolar y el mejoramiento de las prácticas docentes en las instancias de desarrollo profesional docente.

### 1.2.2. Herramientas para apoyar el uso de las evaluaciones

Dado que cada estado propone usos particulares de los resultados dentro del marco señalado anteriormente, a continuación, a modo de ejemplo, se detallan iniciativas del estado federal de Nordrhein-Westfalen según lo propuesto por su Agencia de Calidad Escolar (QUA-LiS NRW).

En su página web QUA-LiS NRW ofrece amplia información para que los profesores puedan usar constructivamente los resultados obtenidos en las pruebas comparativas, teniendo como base el modelo de la figura 1. Las herramientas ofrecidas se agrupan en secciones en el menú de resultados y realimentación de la página web de esta agencia; cada una de ellas funciona como insumo específico para alguna de las etapas que componen el modelo. No existe información acerca de actividades con profesores u otros actores relevantes que estén organizadas desde la agencia de calidad escolar.

**Figura 1** Modelo de uso de resultados sugerido por la agencia de calidad escolar de Nordrhein-Westfalen (QUA-LiS NRW)



Fuente: <http://www.schulentwicklung.nrw.de/vera3/lehrerinformationen/hintergrund/ziele/> (traducción propia).

La información ofrecida es la siguiente:

- Un documento que explica la presentación de resultados de los reportes en sus distintos niveles (individual, clase y colegio). El objetivo es que los profesores puedan leer y entender los resultados obtenidos y lo que significan las distintas representaciones gráficas utilizadas. Estos ejemplos de cada uno de los reportes del estudio indican qué significa cada uno de sus componentes y abreviaturas utilizadas, así como detalles de la información entregada y las diferencias entre representaciones gráficas en términos de dicha información.
- Indicaciones y pistas para una correcta interpretación de los resultados. Esta sección consiste en la presentación de ejemplos de resultados con preguntas que los profesores pueden hacerse con distintos propósitos, por ejemplo:
  - Relacionadas con los niveles de desempeño: ¿existen diferencias marcadas entre los distintos ámbitos evaluados?; ¿cómo le fue a su clase en comparación con el colegio y con el estado?; ¿cómo se explican estas diferencias? Para esto pueden considerarse aspectos tales como: la composición de la clase, las prácticas de aula, los materiales didácticos utilizados y focos temáticos de la prueba.
  - Relacionadas con los contenidos de las pruebas: para alemán, ¿con qué tipo de textos se trabajó comprensión lectora en el aula?, con instrucciones como: "Compare esta información con la tabla de las frecuencias de repuestas correctas de sus estudiantes". Para matemáticas: ¿con qué contextos y enunciados se trabajó la resolución de problemas en el aula? En clases, ¿se trabajó la selección de información de diagramas o la presentación de información actualizada?
  - De orientación general: ¿qué resultados llaman la atención, son inesperados o requieren especialmente una explicación? ¿Qué relación tienen los resultados con otras evaluaciones de los estudiantes o con sus calificaciones escolares? ¿Pueden relacionarse los resultados con elementos de práctica de aula? ¿Qué resultados deben discutirse con los demás profesores especialista de la asignatura?
  - Transversales entre pruebas: ¿existen similitudes entre los niveles de desempeño obtenidos por los estudiantes en comprensión lectora en inglés y en alemán?
- Material para preparar la realimentación a los estudiantes. Aquí se consideran cuestionarios para preguntar a los estudiantes sobre sus impresiones luego de la aplicación y su opinión respecto de los ítems aplicados, así como un instructivo para que el profesor prepare actividades de reflexión a propósito de la prueba, una presentación de Power Point para guiar la actividad, e incluso un video en que se simula una entrevista de realimentación con un estudiante. Es importante destacar que el formato propuesto es una entrevista individual con cada estudiante.
- Instructivos con orientaciones para el mejoramiento de las prácticas de aula: en éstos se incluyen ejemplos concretos de interpretación de resultados de una clase con las correspondientes conclusiones y ejemplos de remedios que pueden llevarse a cabo en aula.

En términos generales, se observa que hay bastante material disponible para trabajar con los resultados de las pruebas comparativas. Si bien el material está bien organizado y es claro, contiene mucha información bastante compleja, lo que supone un usuario familiarizado con este tipo de resultados que además dispone de mucho tiempo para procesar y analizar la información.

Adicionalmente, en lo relativo al desarrollo de las prácticas docentes, llama la atención que todas las estrategias están basadas en un conocimiento específico de aquello que los estudiantes respondieron correcta e incorrectamente de acuerdo con ítems que están a disposición de los profesores. Si bien no se indica explícitamente la liberación de los ítems, sólo parece viable realizar las actividades propuestas contando con ellos. En el mismo sentido, la entrevista individual con los estudiantes requiere un formato muy exigente para el profesor en términos de la preparación, así como estudiantes que estén habituados a ese tipo de intercambios, y a reflexionar y opinar respecto de las actividades pedagógicas en que participan.

### 1.2.3. Uso de los resultados

La información reportada aquí corresponde a un resumen de distintas presentaciones acerca de las pruebas comparativas efectuadas en seminarios de la Asociación de Desarrollo Escolar Basado en Evidencia Empírica (EMSE). Esta asociación está compuesta por 80 representantes de las distintas agencias de calidad escolar de los 16 estados federales, así como de instituciones académicas y universidades.

En términos generales se plantea que la actitud del profesorado respecto de las pruebas comparativas ha sido indiferente, por lo que se hace una valoración discreta. Sin embargo, se han manifestado reparos concretos respecto de la sobrecarga que éstas implican para los profesores, así como dudas respecto de sus beneficios, y se considera que constituyen una iniciativa que implica costos elevados.

En cuanto al uso efectivo de los resultados, si bien la presentación está bien diseñada y se trata de resultados con mucho potencial de uso, se considera que están subutilizados. Específicamente en un estudio en Hessen se detectó un amplio desconocimiento respecto de los materiales didácticos a disposición de los profesores.

En un estudio hecho por la agencia de evaluación del estado de Sarre acerca del uso de los resultados de VERA 8 de francés e inglés en ese estado ( $n=113$ ), se encontró que cerca de 80% de los profesores dice sacar conclusiones de VERA para diseñar sus clases y reflexionar en función de los niveles de desempeño. Sin embargo, sólo 53% de los profesores de inglés y 36% de los de francés consideran importantes los resultados obtenidos (Meyer, 2014).

Según datos recolectados en Sajonia, se ha visto que a propósito de los resultados el hincapié se hace más en el desempeño del estudiante que en su eventual uso como reflexión del docente acerca de su propio trabajo.

De igual manera, existe una percepción relativamente generalizada en Sajonia y Hessen respecto del escaso uso que se da de los resultados para un verdadero mejoramiento de las prácticas docentes. Teniendo esto como antecedente, el estado de Hessen decidió crear la figura del asesor de desarrollo de prácticas de aula, quien va a las escuelas a colaborar en la adquisición de capacidades en torno al uso de resultados de VERA. Esta iniciativa aún debe ser evaluada. En este mismo estado, un grupo de investigadores de la universidad de Kassel (Vogel, Achmetli, Blum y Krawitz, 2014) realizó un estudio experimental en 29 colegios para mejorar el uso de los resultados de las pruebas comparativas. En el grupo experimental se implementaron cuatro instancias de desarrollo profesional a lo largo de un año escolar (dos antes y dos después de la

aplicación de las pruebas comparativas). Los resultados fueron positivos para este grupo en términos de un aumento en la aceptación de las pruebas y la percepción acerca de su utilidad con propósitos diagnósticos, y como insumo para las prácticas docentes. No obstante, se conservó la crítica acerca de la sobrecarga laboral que supone.

#### 1.2.4. Opinión pública

Producto de la búsqueda de opiniones en medios de opinión pública se encontraron más bien críticas generales a la estructura de las pruebas y sus efectos no deseados en los colegios, sin que se aludiera específicamente a los resultados o sus usos.

Por ejemplo en el *Der Spiegel*<sup>3</sup> hay una nota de mayo de 2014 (pocos días después de la aplicación de las pruebas) en la que se critica que las pruebas contribuyen a aumentar la presión por el rendimiento en las escuelas, lo que estresa a los profesores y a los estudiantes. Desde el sindicato de profesores, se criticó también que los malos resultados en escuelas problemáticas no hayan tenido hasta ahora ninguna consecuencia en términos de mayor contratación de profesores o implementación de medidas pedagógicas relevantes. También se criticó la asignación de recursos para acciones estratégicas que no se sabe qué efecto tienen, si alguno tuviera, en tanto no existen estudios independientes al respecto.

En cuanto a la opinión de los padres, también se reportan críticas acerca de que la prueba de tercer grado es excesivamente extensa e incluye contenidos que los estudiantes no han visto en clases, por lo que tendría un efecto motivacional negativo, además de resultar excesivamente agotadora.

En cuanto a las opiniones favorables, se plantean en términos de la importancia de la estandarización de la enseñanza en los distintos estados federales, sin embargo, no se hacen alusiones específicas a los resultados.

Se destaca la prohibición de hacer *rankings* de colegios o los estados con los resultados, y también que a la prueba no se le puede asignar una nota.

#### 1.2.5. Ejemplos de materiales de apoyo

La siguiente página de internet contiene todos ejemplos de informes de resultados descargables, así como videos explicativos y material didáctico complementario:

- <http://www.schulentwicklung.nrw.de/lernstand8/lehrerinformationen/ergebnisrueckmeldung/index.html>

<sup>3</sup> *Der Spiegel* (en español, *El Espejo*), es una revista de noticias semanal que se publica en la ciudad de Hamburgo. Es una de las principales publicaciones de este tipo en Europa y la más importante de Alemania, con un tiraje semanal cercano al millón de ejemplares. Es conocida en Alemania por su estilo académico, así como por su influencia en la opinión pública. Una edición media tiene unas 170 páginas, con dos de contenido por cada una de publicidad.

# 2

## El caso de Chile

---

### 2.1. Características de la evaluación de gran escala

#### 2.1.1. Introducción

El actual sistema nacional de evaluación de resultados de aprendizajes de Chile fue creado en 1988. Este sistema es heredero de una serie de esfuerzos para monitorear los resultados de aprendizaje a nivel nacional, como la Primera Prueba Nacional a Estudiantes de Octavo Básico (1968 a 1971) y el Programa de Evaluación del Rendimiento Escolar (1982 a 1984). Actualmente la evaluación de resultados nacional es conocida como Sistema de Medición de la Calidad de la Educación y es referido usualmente por su acrónimo: SIMCE.

Un hito importante en el desarrollo de las pruebas SIMCE ocurrió en 1998, cuando se pasó del uso de teoría clásica al uso de teoría de respuesta al *ítem*, la cual se utiliza hasta el día de hoy. Este cambio permitió la equiparación de los puntajes año a año, de modo que los resultados de las pruebas SIMCE pueden utilizarse para estudiar tendencias partiendo desde 1998 hasta las pruebas actuales.

En el reporte oficial de la Comisión para el Desarrollo y Uso del Sistema de Medición de la Calidad de la Educación (CDU, de aquí en adelante), convocada en 2003, se describen como los propósitos iniciales del SIMCE:

entregar información para orientar a distintos agentes involucrados en la administración y planificación del sistema educativo. Complementariamente, este sistema de medición buscaba estimar la calidad de cada establecimiento educativo, comparar dichos resultados, buscar los factores explicativos y evaluar los resultados de los programas pedagógicos. Por otra parte, buscaba también orientar la actividad de perfeccionamiento docente, las acciones de supervisión y la asignación de recursos (MINEDUC, 2003: 26).

La aplicación periódica de las pruebas SIMCE y su uso para generar indicadores públicos de resultados de los establecimientos educacionales en Chile se consiguieron durante la dictadura militar en la Ley Orgánica Constitucional de Enseñanza, en 1990.

En su análisis del estado del SIMCE en 2003, la CDU constató que “El SIMCE y sus resultados son ampliamente utilizados en el diseño y evaluación de políticas educativas” (MINEDUC, 2003: 35), y listó diversos usos que se le dan a sus resultados: instrumento de alineamiento curricular, instrumento de información para orientar la demanda, instrumento de focalización, instrumento de evaluación de políticas, indicador de mejora a nivel nacional, e indicador para la asignación de incentivos.

Una evaluación reciente del SIMCE y sus usos realizada por el Equipo de Tarea para la Revisión del SIMCE (MINEDUC, 2015) constata que desde la revisión de 2003 de la CDU ha existido un incremento en las consecuencias asociadas a los resultados de las pruebas. De forma crucial, la ley que estableció el Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización (Ley 20529, de ahora en adelante Ley de Aseguramiento de la Calidad), promulgada en 2011, estableció el uso del SIMCE como parte central de los indicadores a ser utilizados en una estructura de incentivos y sanciones a los establecimientos educacionales.

### 2.1.2. Marco administrativo/legal

El SIMCE estuvo históricamente administrado a nivel nacional por el Ministerio de Educación desde sus inicios en 1988 hasta 2012, año en que el control de la evaluación nacional fue traspasado a la Agencia de Calidad de la Educación —institución independiente del Ministerio de Educación—, creada como parte de la Ley de Aseguramiento de la Calidad con el propósito expreso de “evaluar y orientar el sistema educativo para que éste propenda al mejoramiento de la calidad y equidad de las oportunidades educativas, considerando las particularidades de los distintos niveles y modalidades educativas” (Ley 20529).

De acuerdo a la Ley de Aseguramiento de la Calidad, la agencia de calidad deberá cumplir con las siguientes funciones:

- Evaluar los logros de aprendizaje de los alumnos de acuerdo con el grado de cumplimiento de los estándares referidos a los objetivos generales señalados en la ley y sus respectivas bases curriculares, por medio de instrumentos y procedimientos de medición estandarizados y externos a los establecimientos educacionales. Asimismo, deberá evaluar el grado de cumplimiento de los otros indicadores de calidad educativa.
- Realizar evaluaciones del desempeño de los establecimientos educacionales y sus sostenedores con base en los estándares indicativos de desempeño.
- Ordenar los establecimientos educacionales en función de las mediciones de los resultados de aprendizaje y de los otros indicadores de calidad educativa con la finalidad, entre otras, de identificar, cuando se requiera, las necesidades de apoyo.
- Validar los mecanismos de evaluación de docentes de aula, directivos y técnico-pedagógicos.
- Proporcionar información en materias de su competencia a la comunidad en general y promover su correcto uso.

Específicamente, la ley establece la obligatoriedad de la aplicación de las mediciones a todos los establecimientos educacionales reconocidos oficialmente por el Ministerio de Educación chileno. De acuerdo con la ley, estas mediciones serán periódicas y censales, y cubrirán un curso de enseñanza básica (primaria) y un curso de enseñanza media (secundaria) en múltiples sectores de aprendizaje.

### 2.1.3. Procedimientos y metodología

La construcción de las pruebas está en su totalidad a cargo del Departamento de Construcción de Pruebas (DCP), el cual depende de la División de Evaluación de Logros de Aprendizaje de la Agencia de Calidad de la Educación. En general, el desarrollo de las pruebas consiste de cinco etapas: la definición del marco de referencia y preparación de las especificaciones; la elaboración de ítems; el ensamblaje y la aplicación de pruebas experimentales; el análisis de resultados de las pruebas experimentales; y, finalmente, el ensamblaje de la prueba definitiva. Este proceso de cinco etapas se extiende por alrededor de un año y medio.

Como se indicó previamente, la legislación chilena establece una aplicación de carácter censal. Además de las pruebas censales, existen dos tipos de pruebas de carácter muestral. La primera de éstas la constituyen las pruebas experimentales, que son aplicadas para recolectar datos respecto al comportamiento psicométrico de las pruebas. El segundo tipo de prueba muestral corresponde a pruebas definitivas que son administradas de forma excepcional, tales como la Prueba Nacional de Educación Física, realizada en 2013.

De acuerdo al informe técnico del SIMCE 2013, el plan de muestreo de dicha prueba fue el siguiente (Agencia de la Calidad de la Educación, 2015: 66):

- **Unidad de muestreo:** las aplicaciones muestrales consideraron como unidad de muestreo el establecimiento educacional, debido a que las pruebas son construidas para reportar resultados a un nivel de agregación mínimo de establecimientos y no a nivel de estudiantes.
- **Marco muestral:** corresponde al listado de todos los establecimientos de las regiones correspondientes en funcionamiento que tienen probabilidad de ser escogidos dentro de la muestra, y que cuenten con estudiantes en el grado que debe ser evaluado. Este listado contiene la identificación del establecimiento (código de identificación, nombre, ubicación geográfica), descripción del grado evaluado (número de cursos y estudiantes) y las variables de estratificación.
- **Tipo de muestreo:** un muestreo estratificado que se ajusta adecuadamente a las características de la población educacional chilena, en el cual se utilizan tanto variables explícitas (que definen los estratos) como implícitas (se verifica su distribución una vez seleccionada la muestra).

La muestra es estratificada por regiones geográficas del país (entre las que pueden excluirse sectores de difícil acceso, como Isla de Pascua), y de acuerdo con la dependencia administrativa del establecimiento (municipal, subvencionado y particular pagado). Con respecto a los tamaños muestrales, la prueba muestral 2013 utilizó los siguientes criterios establecidos en el interior de la Agencia de Calidad de la Educación: no sobrepasar los 700 establecimientos, mil cursos y 30 mil estudiantes, ya que esta cantidad se acepta como suficiente para garantizar resultados representativos, con una holgura de 5%.

La aplicación de las pruebas SIMCE presenta un desafío logístico considerable, que puede ser caracterizado en cinco etapas: el empadronamiento o la construcción de los listados de establecimientos educacionales que participaran de las pruebas; la impresión y el mecanizado (por ejemplo, la organización del material en contenedores asociados a cada sitio de aplicación); la distribución; la aplicación; y la entrega para su captura (como el proceso de recolección de materiales aplicados).

De acuerdo con el reporte técnico SIMCE 2013, los manuales de administración, las instrucciones de aplicación y todas las condiciones en que se rinde la prueba se preparan y se aplican de forma consistente para todos los estudiantes. Los procedimientos para resguardar los materiales aplicados son coordinados de manera que el almacenamiento, la distribución y la recuperación de éstos se realicen de forma segura y adecuada.

Para optimizar la calidad de la aplicación y que todos los estudiantes sean evaluados en igualdad de condiciones, existen procesos que la Agencia de Calidad externaliza. Estos procesos se realizan mediante licitaciones de servicios reguladas por bases que contienen tanto las especificaciones técnicas que debe ofrecer el servicio y las tareas definidas sobre éste, como especificaciones de tipo administrativo, por ejemplo, las responsabilidades y multas estipuladas en caso del incumplimiento de los servicios. Cabe destacar que cada contrato considera cláusulas que aseguran la confidencialidad y la seguridad del material, y que involucran a todo el personal que participa en las etapas de impresión, distribución, aplicación y retiro del material, y en el proceso de captura.

Como lo indica el reporte técnico 2013 de las pruebas SIMCE, las pruebas se construyen con miras a ser “representativas del currículo nacional vigente al momento de la aplicación”.

Las pruebas SIMCE comprenden actualmente aplicaciones a los cursos de segundo, cuarto, sexto y octavo básico (educación primaria), y segundo y tercero medio (educación secundaria). En términos generales, el sistema educacional chileno establece que los estudiantes de educación básica tengan seis años de edad al ingresar a primero básico, y terminen primaria con 13 años al cursar octavo básico. Asimismo, se espera que, en general, los alumnos de educación media (secundaria) tengan 14 años cuando cursen primero medio y tengan 17 años cuando cursan cuarto medio.

La figura 2 resume el calendario de aplicación de pruebas SIMCE entre 1998 y 2015 para los distintos niveles y áreas de contenidos.

## 2.2. Características de las prácticas de uso de los resultados de las evaluaciones

### 2.2.1. Introducción

Desde su creación en 1988, el sistema de evaluación ha ido paulatinamente sumando nuevos usos y consecuencias asociadas a los resultados de las pruebas SIMCE. Por ejemplo, en 1996 las pruebas fueron incluidas como parte de un sistema de incentivos para el desempeño de los establecimientos educacionales —Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados (SNED)— y en 2011 la Ley de Aseguramiento de la Calidad de la Educación estableció el uso de las pruebas SIMCE como un indicador central en la ordenación de

**Figura 2** Calendario de aplicaciones por curso y área curricular

Grado	Área	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015		
2° básico	Lenguaje y Comunicación: Comprensión de Lectura																				
	Lenguaje y Comunicación: Comprensión de Lectura																				
	Matemática																				
	Comprensión del Medio Natural, Social y Cultural																				
4° básico	Ciencias Naturales																				
	Historia, Geografía y Ciencias Sociales																				
	Discapacidad Sensorial																				
6° básico	Lenguaje y Comunicación: Comprensión de Lectura																				
	Lenguaje y Comunicación: Escritura																				
	Matemática																				
	Ciencias Naturales																				
8° básico	Historia, Geografía y Ciencias Sociales																				
	Discapacidad Sensorial																				
	Lenguaje y Comunicación: Comprensión de Lectura																				
	Matemática																				
II medio	Ciencias Naturales																				
	Historia, Geografía y Ciencias Sociales																				
	TIC																				
	Inglés																				

Fuente: <http://simce.cl/>

los establecimientos educacionales, así como la asignación de castigos e incentivos con base en el desempeño de los establecimientos.

Como se mencionó anteriormente, en su análisis del estado del SIMCE en 2003, la CDU constató que “El SIMCE y sus resultados son ampliamente utilizados en el diseño y evaluación de políticas educativas” (MINEDUC, 2003: 35), y listó distintos usos que se le dan a los resultados producidos por el SIMCE:

- **Instrumento de alineamiento curricular:** las pruebas dan una clara señal a los colegios y profesores respecto a la necesidad de adhesión al currículo.
- **Instrumento de información para orientar la demanda:** los resultados del SIMCE se entregan con el fin de informar a los padres y apoderados respecto a la calidad de los establecimientos, en el marco de un modelo de mercado que, en teoría, promovería la migración de los estudiantes desde colegios con peores resultados a colegios con mejores resultados en SIMCE.
- **Instrumento de focalización:** los resultados de las pruebas sirven para identificar establecimientos y sectores que podrían beneficiarse de intervenciones y programas de apoyo.
- **Instrumento de evaluación:** los resultados del SIMCE se utilizan para estimar el efecto de intervenciones en todo nivel del sistema, desde intervenciones y programas internos de un establecimiento hasta iniciativas a escala nacional.
- **Indicador de mejora a nivel nacional:** los resultados de las pruebas son utilizados para la fijación y el monitoreo de metas a nivel nacional sobre el progreso del sistema educacional.
- **Indicador para la asignación de incentivos:** los resultados de las pruebas son utilizados como un indicador central del SNED, el cual identifica establecimientos de alto desempeño y premia a los profesores de dichos establecimientos con bonos salariales.

Si bien la CDU valoró estos usos a nivel de políticas educacionales, el reporte emitido indicó que el uso a nivel de política pública prevalecía sobre el uso pedagógico y de gestión curricular en el interior de los establecimientos educacionales, por las características de la información entregada y por las brechas de preparación de docentes en el ámbito de la evaluación de aprendizajes. También se diagnosticó un escaso uso y conocimiento por parte de los apoderados y padres de familia, lo cual se atribuyó en ese momento a la insuficiencia de los canales de comunicación y a la falta de pertinencia de la información proporcionada para los requerimientos de los padres.

Al referirse a la comunicación pública de los resultados nacionales, la comisión estableció que ésta se concentraba en los logros de aprendizaje sin reportar, junto con otros indicadores de la educación (como cobertura o tasas de retención), lo que hubiese favorecido una visión más completa de la situación de la educación nacional.

Asimismo, la comisión recibió testimonios de consecuencias no deseadas, como las prácticas de exclusión y selección de estudiantes con bajo rendimiento como estrategia para mejorar los resultados del SIMCE, y la tendencia de los establecimientos a concentrar sus mejores recursos y esfuerzos en las generaciones y asignaturas evaluadas, en desmedro de las demás (MINEDUC, 2015: 12).

Con base en las recomendaciones de la CDU, se realizó una serie de cambios que incluyeron mejoras a los reportes de resultados y la capacitación de profesores y directivos de establecimientos en la interpretación de los resultados de las pruebas.

## 2.2.2. Herramientas para apoyar el uso de las evaluaciones

Además de ser la encargada del diseño y la aplicación de las pruebas SIMCE, la Agencia de Calidad de la Educación es también la institución responsable de su análisis y comunicación de los resultados. De acuerdo con la información que la agencia proporciona en su sitio web, los resultados de las pruebas SIMCE son entregados por tres medios principales: reportes para docentes y directivos para cada curso que participó en la prueba (entregados a cada establecimiento); reportes para padres y apoderados sobre cada nivel evaluado por las pruebas, y en el sitio web de la Agencia de la Calidad de la Educación. De manera adicional, la agencia ofrece presentaciones oficiales a la prensa para dar a conocer los resultados anuales de las pruebas; pone a disposición del público informes técnicos sobre el SIMCE, y da acceso a las bases de datos con los puntajes de los establecimientos de todo el país.

El sitio web de las pruebas SIMCE ([www.simce.cl](http://www.simce.cl)) sirve como repositorio central para acceder a información relacionada con esta evaluación. Como se muestra en la figura 3, el sitio está organizado en términos generales con base en una sección de información general respecto a las pruebas y una serie de secciones dedicadas a cuatro grupos de usuarios: padres, apoderados y estudiantes; sostenedores y directores; docentes y asistentes, e investigadores.

Figura 3 Extracto de la página [www.simce.cl](http://www.simce.cl)

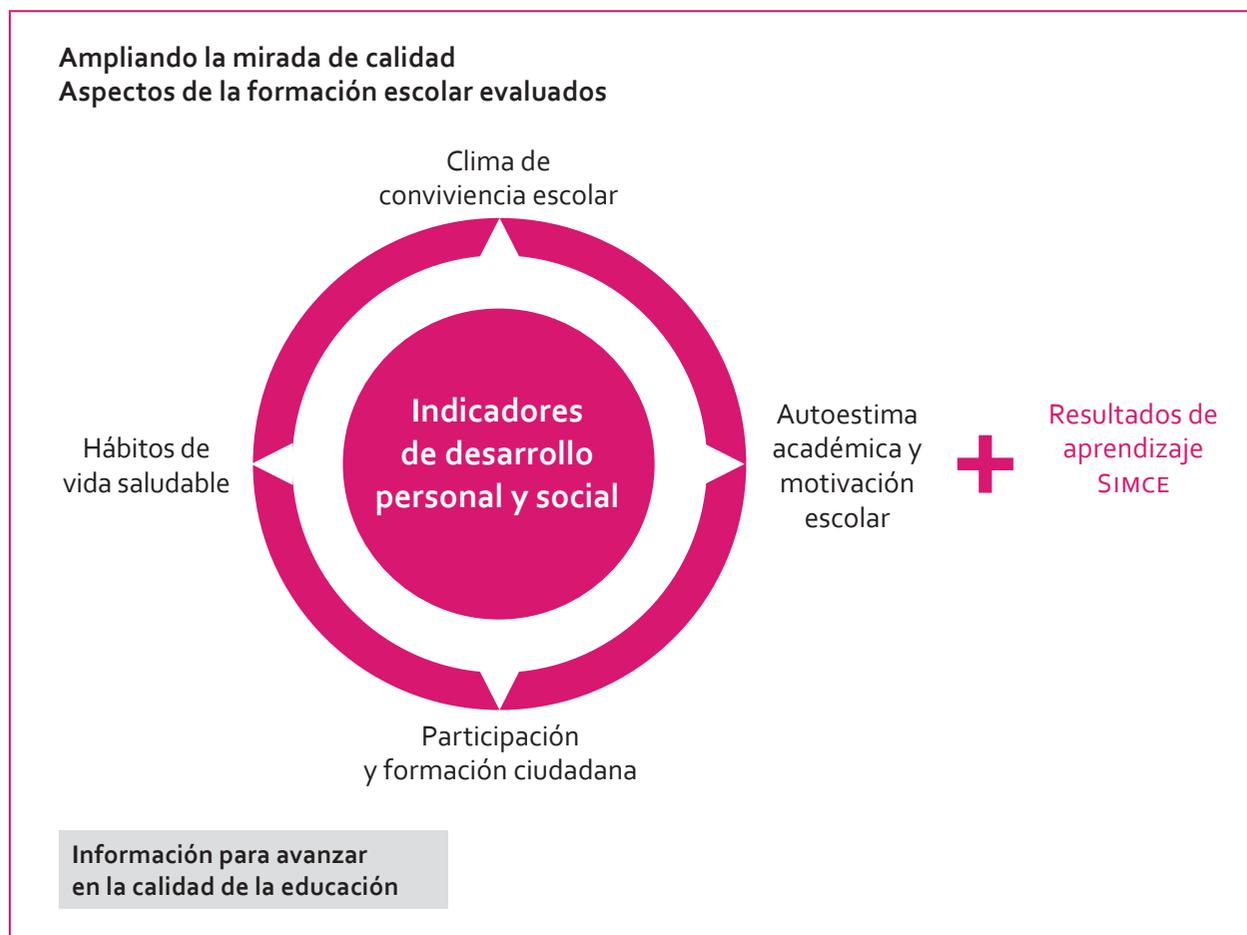
The screenshot shows the website interface for the Agencia de Calidad de la Educación. At the top left is the logo and name of the agency, along with 'Gobierno de Chile'. A navigation bar includes links for 'CONTACTO', 'PREGUNTAS FRECUENTES', 'MAPA DEL SITIO', and 'LOGIN'. A search bar is located on the right with the text '¿Qué buscas?'. Below the navigation bar is a horizontal menu with categories: 'NOSOTROS', 'NOTICIAS', 'SIMCE', 'COORDINACIÓN SAC', 'VISITAS DE EVALUACIÓN Y ORIENTACIÓN', and 'ESTUDIOS INTERNACIONALES E INVESTIGACIONES'. The main content area features a 'PREGUNTAS FRECUENTES' section with a list of user groups: 'Participa de la mejora educativa', 'Padres, Apoderados y Estudiantes', 'Sostenedores y Directores', 'Docentes y Asistentes', and 'Investigadores'. Below this is a 'Buscador de resultados (Simce)' section with a text input field labeled 'Escriba Nombre o RBD aquí'. To the right of the user groups, there is a text block titled '¿Para qué sirven los resultados de las pruebas Simce?' which explains that the results provide information on learning standards and complement the analysis performed by each establishment based on its own evaluations.

Fuente: <http://www.agenciaeducacion.cl/preguntas-frecuentes/#simce>

La sección de información general en la página web proporciona acceso a información de difusión respecto a qué es el SIMCE, sus usos, calendarios de aplicación, resultados de las pruebas y preguntas frecuentes. De forma adicional, la página presenta un buscador que permite obtener resultados de cada establecimiento en Chile según su nombre o su código identificador. La página permite acceder con facilidad a la información pública disponible respecto a las pruebas SIMCE.

Los resultados del SIMCE se comunican de forma anual en una o más conferencias de prensa mediante las cuales la Agencia de Calidad de la Educación presenta un resumen de los principales resultados de las pruebas aplicadas en el año correspondiente. Estas conferencias de prensa se realizan a niveles tanto nacional como regional. El contenido de las conferencias se centra en la exposición de tendencias y resultados a nivel nacional, y se presentan conclusiones generales respecto a la tendencia de los resultados en relación con años anteriores. Esta información de resultados tiene asociado un documento de presentación que se hace público a través de la página web del SIMCE. Cabe destacar que en años recientes estas presentaciones incluyen también un reporte de indicadores de desarrollo personal y social de los estudiantes (ver figura 4), a fin de contextualizar los resultados de aprendizajes asociados a las pruebas.

**Figura 4** Diagrama de indicadores que son presentados a la prensa



Fuente: <http://simce.cl/>

Los resultados de las pruebas SIMCE también se dan a conocer mediante reportes a padres y apoderados. Estos reportes son entregados a cada establecimiento para ser repartidos entre los padres y apoderados de los estudiantes. De forma adicional a estas copias en papel que se entregan por medio del establecimiento, la página web del SIMCE permite descargar los reportes de cualquier establecimiento del país mediante una búsqueda por su nombre o por su código identificador, lo que le permite a cada padre o apoderado obtener el reporte específico del establecimiento al que asiste cada estudiante.

El reporte SIMCE 2014 de octavo básico para padres y apoderados indica en su introducción:

Estamos convencidos, y así lo demuestra la evidencia, que el apoyo y preocupación de los padres y apoderados de cada niño son muy importantes en su educación y en lo que puedan aprender. La calidad de la educación la construimos entre todos y requiere de nuestro compromiso el asegurar que todos los estudiantes tengan las herramientas para cumplir sus sueños y, desde distintos ámbitos, ser un aporte a la sociedad.

El documento que hoy le hacemos llegar es muy importante, pues queremos que usted esté informado sobre los aprendizajes de los jóvenes de su establecimiento, de tal manera que usted pueda participar activamente en el proceso escolar de su hija e hijo, y avanzar juntos en mejorar la calidad de la educación.

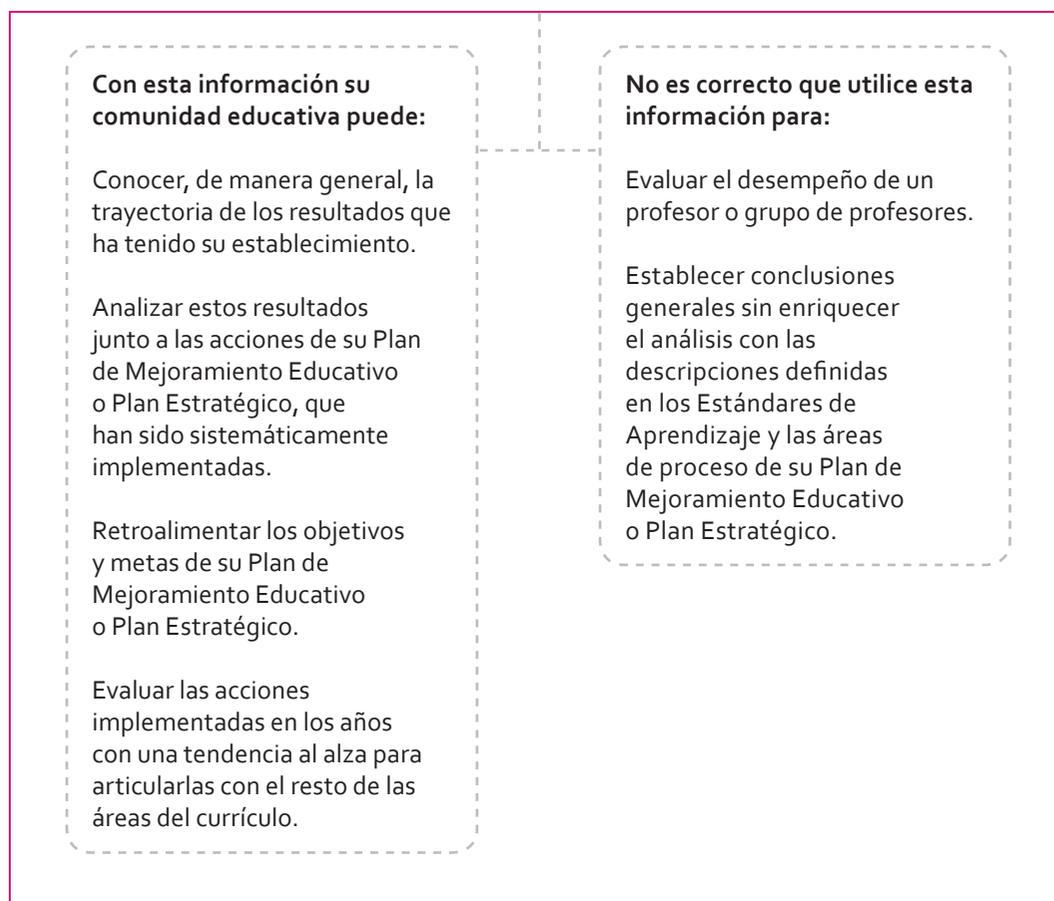
Se trata de un reporte breve (de aproximadamente ocho páginas) que se divide en cuatro partes principales: la presentación y la explicación de los tres niveles de desempeño utilizados para clasificar a los estudiantes según sus resultados en las pruebas del SIMCE (Insuficiente, Elemental y Adecuado); los porcentajes de estudiantes en cada nivel de desempeño dentro del establecimiento educacional; información respecto a la ubicación relativa de los resultados del establecimiento en relación con otros similares (por ejemplo, de condiciones socioeconómicas similares), y, finalmente, un conjunto de recomendaciones de posibles formas en que los padres y apoderados pueden apoyar el aprendizaje de los estudiantes.

Por su parte, los reportes a profesores y directivos se entregan directamente a los establecimientos por la Agencia de Calidad de la Educación y se encuentran disponibles también a través de la página web del SIMCE ([www.simce.cl](http://www.simce.cl)). Los reportes de resultados a profesores y directores son considerablemente más detallados que los reportes a padres y apoderados, y están organizados en tres capítulos principales: los resultados de aprendizaje del establecimiento, las orientaciones para los equipos directivos respecto a la gestión escolar, y las orientaciones a los docentes respecto a la gestión pedagógica.

El primer capítulo, sobre los resultados de aprendizaje, contiene:

- Los puntajes promedio de los estudiantes de cada nivel que participó en la prueba, y las comparaciones con pruebas anteriores.
- Los porcentajes de estudiantes que se encuentran en cada uno de los tres niveles de desempeño definidos en los estándares nacionales de aprendizaje (Insuficiente, Elemental y Adecuado), y la evolución de estos porcentajes en el tiempo.
- Los puntajes por curso y su comparación con establecimientos de similar perfil socioeconómico.

**Figura 5** Ejemplo de las recomendaciones respecto a cómo usar y como no usar los resultados reportados



Fuente: <http://simce.cl/>

Los resultados en este reporte van acompañados por recomendaciones explícitas sobre qué usos son recomendados y qué usos o interpretaciones no se consideran correctas (ver figura 5).

El segundo capítulo, dirigido a los directivos del establecimiento, se concentra en dar orientaciones para optimizar la gestión escolar con el fin de lograr mejores aprendizajes. En la introducción de este capítulo, en el reporte de resultados 2014, se indica:

En este capítulo entregamos a usted y a su equipo directivo algunas orientaciones que guían, junto a otras fuentes de información, el análisis e interpretación de los resultados entregados en el capítulo anterior. El propósito es que identifique y reflexione sobre las acciones o estrategias que han permitido mejoras en los aprendizajes y aquellas que requieren ser modificadas, con el fin de fortalecer los procesos y capacidades internas, y de ese modo, avanzar en el cumplimiento de los desafíos institucionales.

Las orientaciones son presentadas en el marco de una propuesta de plan de trabajo con cuatro etapas: diagnóstico, planificación, implementación/monitoreo y evaluación.

Finalmente, el tercer y último capítulo se concentra en la gestión pedagógica, y tiene como objetivo (de acuerdo con los reportes 2014) orientar las actividades de análisis de los resultados de aprendizaje SIMCE y de las evaluaciones de aula (trabajos de investigación, presentaciones, pruebas orales y escritas, portafolios, entre otros), para que este análisis se transforme en una práctica constante y útil que permita fortalecer los procesos y favorezca la mejora de los aprendizajes de sus estudiantes.

Este capítulo está organizado en torno a un modelo de interpretación y uso de los resultados de evaluación de cuatro etapas: el análisis y la interpretación de resultados; el diseño de propuestas para mejorar aprendizajes, implementación y monitoreo, y, finalmente, la evaluación de logros.

Como parte de la información que se proporciona para promover la reflexión respecto a la gestión pedagógica, se incluyen en el reporte ejemplos de preguntas utilizadas en las pruebas. A cada pregunta liberada la acompañan un análisis de su relación con el currículo y los estándares de aprendizaje y un comentario respecto a sus características específicas y sus opciones de respuesta.

Otra importante herramienta para la difusión y la utilización de la información generada por el SIMCE es la publicación de las bases de datos. Se puede acceder a las bases de datos con los resultados a nivel de establecimientos educacionales, comunas y regiones libremente a través del sitio web del SIMCE ([www.simce.cl](http://www.simce.cl)). Estos datos están disponibles para todas las pruebas a partir de 1998. La Agencia de Calidad de la Educación ofrece a la comunidad académica una serie de bases de datos adicionales con identificadores a nivel de estudiante; sin embargo, estas bases son de carácter restringido, y para acceder a ellas es necesario presentar una solicitud que incluya una propuesta de investigación y comprometerse a respetar ciertas condiciones para asegurar el uso ético y respetuoso de los datos (por ejemplo, la confidencialidad de la información).

Finalmente, la Agencia de Calidad de la Educación produce informes técnicos anuales que detallan el proceso de diseño y construcción de las pruebas, las operaciones logísticas de campo, el procesamiento y el análisis de datos, y la comunicación de los resultados.

### 2.2.3. Uso de los resultados

Con el propósito de “hacer un diagnóstico sobre el sistema y la información que éste proporciona, además de elaborar recomendaciones y propuestas para el desarrollo y uso de la medición tendientes a optimizar su impacto en la calidad de los aprendizajes”, se creó la Comisión para el Desarrollo y Uso del SIMCE (Meckes, 2003: 165-166).

Posteriormente, en 2014 se convocó a una nueva comisión con el nombre de Equipo de Tarea para la Revisión del SIMCE. Esta convocatoria se dio en el contexto de diversos cambios en las políticas públicas de evaluación educacional (como la clasificación de escuelas acorde a la Ley de Aseguramiento de la Calidad) y una serie de “manifestaciones críticas al SIMCE por parte de movimientos sociales y de un grupo de investigadores y académicos que firmaron una carta abierta cuestionando varias características de la medición nacional, como también sus políticas asociadas” (MINEDUC, 2015: 4).

Ambos reportes examinan en dos momentos distintos los usos intencionados y no intencionados de la aplicación y la comunicación de resultados de las pruebas SIMCE, con base en las cuales

generan recomendaciones para su mejora. En su análisis, el Equipo de Tarea SIMCE consideró explícitamente la necesidad de

hacerse cargo de los usos e interpretaciones de la información que entrega. Para esto, es vital que el modo de presentar la información no induzca usos e interpretaciones erróneas y permita diversas lecturas sobre la calidad de la educación. Además, es imprescindible monitorear que los efectos de la evaluación que se dan en la práctica sean efectivamente los deseados o buscados (MINEDUC, 2015: 1).

El estudio de Taut, Cortés, Sebastian y Preiss (2009) examinó los reportes para directivos/profesores y para apoderados en términos del acceso, la comprensión y el uso de éstos. Los resultados de este estudio mostraron que los profesores y los directivos no tienen problemas en acceder a los resultados, pero que los padres pueden tener dificultades. Por ejemplo, el estudio encontró que los establecimientos donde los resultados fueron positivos tendían a comunicar dichos resultados a todos los padres, pero aquellos con resultados negativos tendían a comunicarlos sólo a los padres de los estudiantes que habían participado en las pruebas. En términos de la comprensión de los resultados, el estudio encontró que la mayoría de los profesores y padres cometían errores en la interpretación de los reportes. Más específicamente, el estudio reveló problemas sistemáticos entre los profesores y directores en la interpretación de cambios/tendencias de los resultados en el tiempo, y problemas para entender el concepto de significancia estadística en las diferencias de resultados.

Asimismo, este estudio reveló que el uso de los reportes no se limita a las conductas esperadas oficialmente, y que existen usos no intencionados entre los participantes del estudio. Por ejemplo, algunos directores reportaron que cada año asignaban a sus mejores profesores a los cursos a los que les correspondía la aplicación de las pruebas SIMCE. Otros directores reconocieron que recompensaban a los profesores cuyos cursos obtenían el mejor promedio en la prueba (lo que contradice explícitamente las recomendaciones de la Agencia de Calidad de la Educación).

Un segundo estudio (Manzi, Bogolasky, Gutiérrez, Grau y Volante, 2014) se interesó en la comprensión y el uso de los resultados SIMCE por parte de los directivos de establecimientos subvencionados. Este estudio encontró, de forma consistente con el estudio de Taut *et al.* (2009), que “aunque se constata un conocimiento mayoritariamente adecuado acerca del SIMCE, subsisten concepciones erradas y preparación insuficiente para la interpretación de los datos” entre los directores de estos establecimientos.

#### 2.2.4. Opinión pública

Los resultados del SIMCE, comunicados a través de por lo menos una conferencia de prensa anual, tienen un gran impacto mediático, y normalmente los cubren de forma prominente tanto la prensa escrita como la radio y la televisión. Sin embargo, esta cobertura se orienta de forma importante a la publicación y la discusión de *rankings* que ordenan los colegios del país de mayor a menor puntaje (ver figura 6).

**Figura 6** Artículo de un periódico nacional publicando con *ranking* de colegios acorde a puntajes de las pruebas SIMCE de segundo básico

*Ranking*: Los 100 colegios con mejor puntaje en SIMCE de 2° básico.

La Agencia de Calidad de la Educación entregó los resultados de la prueba que midió la comprensión lectora a 218 mil alumnos de todo el país. En la lista se consideraron sólo los establecimientos con más de seis alumnos.

<i>Ranking</i>	Colegio	Comuna	Ruralidad	Dependencia	Número de alumnos	Puntaje
1	Escuela Básica Particular Trupán	Conchalí	Urbano	Subvencionado	14	320
2	Escuela Básica Minas De Leuque	Treguaco	Rural	Municipalizado	7	318
3	Escuela Huemul	Teno	Rural	Municipalizado	9	311
4	Escuela Rural Juan A. Vera Díaz	Quelén	Rural	Municipalizado	7	310
5	Escuela Básica Madrigal	La Reina	Urbano	P. Pagado	29	309
6	Escuela Irene Morales	San Javier	Rural	Municipalizado	9	308
7	Colegio Villa María Academy	Las Condes	Urbano	P. Pagado	111	307
8	Escuela Particular Dollimco	Cholchol	Rural	Subvencionado	10	306
9	Colegio Franco Inglés	Viña Del Mar	Urbano	P. Pagado	9	304
10	Escuela Municipal Santa Amalia	Requinoa	Rural	Municipalizado	6	304
11	C.educ.mons.Manuel Larrain Errázuriz	Curicó	Urbano	P. Pagado	22	304
12	Colegio Mauricio Rugendas	La Florida	Urbano	Subvencionado	10	304
13	Colegio Saint George's College	Vitacura	Urbano	P. Pagado	185	303
14	Escuela Marta Colvin	Copiapó	Urbano	Subvencionado	30	303
15	Colegio San José	Angol	Urbano	P. Pagado	28	303
16	Colegio Internacional Sek Del Pacífico	Concón	Urbano	P. Pagado	23	303
17	Escuela Básica La Granja	Cañete	Urbano	Municipalizado	11	303
18	Colegio Nueva Concepción	Temuco	Urbano	Subvencionado	28	302
19	Colegio Constitución	Constitución	Urbano	P. Pagado	21	302
20	Colegio Particular Playground	La Florida	Urbano	P. Pagado	22	302
21	Colegio Padre Hurtado y Juanita De Los Andes	Las Condes	Urbano	P. Pagado	104	302
22	Colegio Santiago College	Lo Barnechea	Urbano	P. Pagado	128	301
23	Colegio Británico St. Margaret's S.A.	Concón	Urbano	P. Pagado	69	301
24	Escuela Zemita	Niquén	Rural	Municipalizado	8	301
25	Colegio The Kent School	Providencia	Urbano	P. Pagado	61	300

Fuente: <http://www.emol.com/noticias/nacional/2013/06/20/604715/ranking-los-100-colegios-con-mejores-puntajes-en-el-simce-de-2-basico.html>

La visibilidad de los resultados SIMCE y su uso para la generación de *rankings*, sumados a los múltiples usos que se le dan a los resultados de acuerdo con la legislación vigente, además de los usos no intencionados (como la preparación de los alumnos específicamente para presentar la prueba y la atribución de responsabilidad directa a los profesores) han dado lugar a un malestar entre un grupo de la opinión pública, el cual se ha consolidado en torno al movimiento Alto al SIMCE (<http://www.altoalsimce.org/>).

Un grupo de académicos, profesores y estudiantes que se han reunido bajo la consigna de “Alto al SIMCE”, justamente para enfrentar este sistema de medición. Se encontraron muchas investigaciones nacionales e internacionales que apuntan a las nefastas consecuencias de la estandarización educativa y al daño que se causa a niños, niñas, jóvenes y profesores (<http://diario.latercera.com/2013/09/07/01/contenido/opinion/11-145773-9-alto-al-simce.shtml>).

Este movimiento ha sido activo en la comunicación de su agenda por medio de redes sociales (Twitter: @AltoalSIMCE, Facebook: <https://es-la.facebook.com/altoalsimce>), periódicos tradicionales, eventos que son cubiertos por estos últimos, o cartas autoridades educativas y columnas de opinión. Alto al SIMCE ha generado, por ejemplo, una serie de videos para comunicar su rechazo al sistema de evaluación en general y las pruebas SIMCE en particular disponibles en el canal de YouTube del movimiento (<https://www.youtube.com/channel/UC4eJh0dqLLtvd9LWSgQIH4g>). Estos videos, de alrededor de diez minutos, tienen la finalidad de explicar las motivaciones del movimiento. También existen videos más cortos (de menos de un minuto) orientados a ser divulgados de forma viral en redes sociales. Estos videos contienen expresiones concisas del rechazo al SIMCE.

De forma más concreta, el movimiento ha promovido el rechazo a las pruebas por parte de los estudiantes y padres instándolos a abstenerse de responderlas, para lo cual se da a los apoderados una plantilla con un “justificativo” para explicar la inasistencia de los estudiantes en los días en los que se aplican las pruebas ([http://www.altoalsimce.org/wp-content/uploads/2015/09/Justificativo\\_Alto-al-SIMCE-2015.pdf](http://www.altoalsimce.org/wp-content/uploads/2015/09/Justificativo_Alto-al-SIMCE-2015.pdf)). Existen otras organizaciones de estudiantes y de padres de familia que se han sumado a los esfuerzos de Alto al SIMCE (ver, por ejemplo, <http://www.publimetro.cl/nota/cronica/secundarios-hacen-llamado-a-no-rendir-el-simce/xlQojz!e2NpIMAl0F7ml/>).

### 2.2.5. Ejemplos de materiales de apoyo

En este caso, los ejemplos de diversos materiales de apoyo utilizados en Chile son los referidos en el apartado anterior.

# 3

## El caso de España

---

### 3.1. Características de la evaluación de gran escala

#### 3.1.1. Introducción

En España la evaluación de tercero de educación primaria comenzó a aplicarse en el curso 2014-2015, y desde entonces se realiza anualmente. A partir del curso 2015-2016 por primera vez se aplicará también al final de la etapa de primaria. Esta evaluación tiene los siguientes antecedentes:

- La evaluación censal de cuarto de educación primaria. Esta evaluación comenzó en el curso 2009-2010 y se aplicó de forma ininterrumpida hasta que fue sustituida por la actual de tercero de primaria.
- La evaluación general de diagnóstico de cuarto de primaria que se llevó a cabo en el curso 2008-2009. El informe completo de esta evaluación se puede consultar en la siguiente dirección: <http://www.mecd.gob.es/dctm/ievaluacion/evaluaciongeneraldiagnostico/pdf-completo-informe-egd-2009.pdf?documentId=0901e72b8015e34e>

Según aparece en el marco general de la evaluación de tercero de educación primaria, la legislación educativa española establece la realización de una evaluación en tercer curso de la educación primaria indicando que “los centros docentes realizarán una evaluación individualizada a todos los alumnos y alumnas al finalizar dicho curso de Educación Primaria”. Se desarrollará según dispongan las administraciones educativas y “comprobará el grado de dominio de las destrezas, capacidades y habilidades en expresión y comprensión oral y escrita, cálculo y resolución de problemas en relación con el grado de adquisición de la competencia en comunicación lingüística y de la competencia matemática”.

Esta evaluación, de carácter diagnóstico, en caso de resultar desfavorable, da al equipo docente la libertad para adoptar las medidas ordinarias o extraordinarias más adecuadas.

La finalidad principal de la evaluación de tercer curso de educación primaria es ofrecer un informe individualizado a cada uno de los alumnos y alumnas sobre su progreso en el grado de adquisición de las competencias en comunicación lingüística y matemática, por lo que permite detectar de manera precoz dificultades en el aprendizaje. Además, se pretende informar a las familias, los

centros y la administración educativa sobre el progreso de los alumnos, y permitir el análisis de los resultados. Finalmente, esta evaluación facilita el establecimiento de medidas de mejora por parte de equipos docentes, directivos, inspección y administración educativa.

Por tanto, los resultados de la evaluación se difunden adecuadamente entre las familias, al profesorado, los centros educativos y la administración educativa.

En un plazo aproximado de siete semanas desde la realización de las pruebas, se presentan un primer informe individualizado de cada alumno o alumna que ofrecerá puntuaciones en cada competencia y en sus respectivas destrezas, y un informe inicial del centro educativo.

Una vez realizadas las valoraciones del Índice Social, Económico y Cultural, se elaboran informes definitivos de los centros educativos que se entregan durante el comienzo del curso siguiente al de la aplicación de la prueba (septiembre-octubre). Estos informes definitivos ofrecen información útil para la reflexión y la toma de decisiones orientadas a la mejora de la educación.

### 3.1.2. Marco administrativo/legal

El marco legal y administrativo está claramente definido y se explica de manera concisa en el marco general de la evaluación del tercer curso de educación primaria:

Como ya se ha señalado, en el artículo 20.3 de la Ley Orgánica 2/2006, del 3 de mayo, de Educación (LOE), modificada por la Ley Orgánica 8/2013, del 9 de diciembre, para la Mejora de la Calidad Educativa (LOMCE), se establece la realización de una evaluación en tercer curso de la educación primaria.

La ley indica además que "las Administraciones educativas podrán establecer otras evaluaciones con fines de diagnóstico" (artículo 144.2), y que "las autoridades educativas establecerán las medidas más adecuadas para que las condiciones de realización de las evaluaciones individualizadas se adapten a las necesidades del alumnado con necesidades educativas especiales" (artículo 144.3).

El Real Decreto 126/2014, del 28 de febrero, por el que se establece el currículo básico de educación primaria, es la norma básica a partir de la cual se desarrollan los cuadros de relaciones de cada una de las competencias evaluadas, cuadros que relacionan los bloques de contenidos y los procesos cognitivos mediante una selección de estándares de aprendizaje aplicables a las evaluaciones externas previstas para educación primaria. Además de esta legislación de carácter estatal, para la evaluación de tercer curso de educación primaria, la referencia legal es la propia de cada administración educativa; en los cuadros de relaciones se podrán tener en cuenta los estándares de aprendizaje evaluables o los criterios de evaluación propios del currículo de cada administración educativa.

En resumen, la legislación estatal establece que todos los alumnos de tercero de primaria tendrán que disponer de un informe que les indicará el progreso en la adquisición de sus competencias, y que serán las administraciones educativas de cada una de las 17 comunidades autónomas quienes indicarán el modo de realizar este informe. Por tanto, habría que acudir a las resoluciones de cada comunidad autónoma para conocer de manera concreta el desarrollo de una ley promulgada por el Ministerio de Educación, Cultura y Deporte (MECD) de España.

### 3.1.3. Procedimientos y metodología

El objeto de evaluación en tercero de primaria está claramente identificado: las competencias en comunicación lingüística y la competencia matemática. Además de la información proporcionada referente a dichas competencias, se debe desagregar en las siguientes destrezas: comprensión y expresión oral y escrita en la primera, y cálculo y resolución de problemas en la segunda.

Como ya se ha dicho, la evaluación es censal; todos los alumnos deben recibir la información sobre la medida en que están adquiriendo las competencias objeto de esta evaluación. Como consecuencia, las pruebas se administran en todos los centros educativos que imparten educación primaria, siendo la población objeto de evaluación los alumnos matriculados en tercero de educación primaria.

La evaluación requiere del empleo de instrumentos con unidades de evaluación adecuadas al tipo de competencias consideradas, que tengan en cuenta situaciones y contextos concretos para que los sujetos puedan demostrar su dominio y aplicación, y cuya administración resulte viable.

Para llevar a cabo esta evaluación se ha decidido emplear pruebas en las que se combinen diferentes formatos de ítems, para conformar instrumentos basados en situaciones-problema que configuran unidades de evaluación. Se trata de pruebas constituidas a partir de casos que sirven como base para la interrogación, y que, en la medida de lo posible, remiten a situaciones similares a las que el alumnado puede encontrar en su vida personal o social. De ese modo, cuando un alumno se enfrenta a una situación-problema está aplicando lo que sabe o sabe hacer a una realidad que le es familiar, o al menos que le resulta verosímil.

Las preguntas formuladas a partir de cada situación-problema van dirigidas, en líneas generales, a comprobar el grado de dominio de las destrezas, capacidades y habilidades que son objeto de la evaluación.

Para el diseño y la elaboración de las unidades de evaluación y de redacción de los ítems se recurre a la colaboración de un grupo de expertos en los ámbitos de la competencia y las etapas evaluadas, incluyendo profesorado de educación primaria con experiencia en la construcción de pruebas con enfoque competencial.

Las siguientes unidades de la Secretaría de Estado de Educación, Formación Profesional y Universidades participan en la evaluación:

- El Instituto Nacional de Evaluación Educativa (INEE) se responsabiliza de la elaboración de las pruebas y los cuestionarios de contexto, así como de los análisis de los resultados de la evaluación y de la elaboración de los informes.
- La Subdirección General de Cooperación Territorial coordina todas las actuaciones de las Direcciones Provinciales de Educación de Ceuta y Melilla relativas a la evaluación.
- La Subdirección General de Promoción Exterior Educativa se responsabiliza de la organización y del desarrollo de la evaluación en los centros en el exterior.

- La Subdirección General de Inspección coordina y supervisa la aplicación y la corrección de las pruebas, y hace el seguimiento de la difusión de los resultados, el análisis y la adopción de medidas en el plan de mejora por los centros.
- Los centros educativos entregan los informes de resultados a los padres, madres o tutores legales de los alumnos y alumnas, los difunden entre la comunidad educativa, los analizan, y adoptan medidas para su mejora.

Las pruebas se realizan en papel y en formato digital. Para evaluar la comprensión oral, se utilizan registros de audio que se reproducen en el momento de la aplicación. La información sobre datos de contexto, de procesos y de recursos se obtiene mediante cuestionarios de contexto. Esta aplicación se lleva a cabo por el profesorado del centro que no imparte clase directa al alumno evaluado.

Las pruebas de la competencia matemática (cálculo y resolución de problemas) y de la competencia en comunicación lingüística (comprensión oral y escrita, y expresión escrita) se realizan en dos días. Las correspondientes a las destrezas de cálculo y resolución de problemas se desarrollan en dos sesiones de 40 minutos cada una; la de expresión escrita en una sesión de 40 minutos, y la de comprensión oral y escrita en una sesión de 50 minutos que comienza con la reproducción de los registros de audio. Entre las diferentes sesiones hay un descanso de 20 minutos.

La evaluación de la expresión oral se realiza a través de entrevistas individualizadas a lo largo de una semana. Las pruebas se aplican durante el mes de mayo del año en curso.

## 3.2. Características de las prácticas de uso de los resultados de las evaluaciones

### 3.2.1. Introducción

Como se ha dicho anteriormente, el antecedente histórico de la evaluación de tercero de educación primaria es la evaluación censal de cuarto de primaria. Ya en estas evaluaciones se presentaba un informe de alumno, otro del centro y un último resumen para la administración educativa competente en la acción de política en materia de educación. Adicionalmente, como consecuencia de la participación de España en distintas evaluaciones internacionales, a partir de 2012 se comenzaron a elaborar y entregar informes con los resultados promedio obtenidos por los centros participantes en estos estudios. Cabe destacar tres informes:

- Informe de centro de los participantes en el Trends in International Mathematics and Science Study (TIMSS) 2011 cuarto grado, 151 centros educativos participantes.
- Informe de centro de los participantes en el Progress in International Reading Literacy Study (PIRLS) 2011 cuarto grado, 312 centros educativos participantes.
- Informe de centro de los participantes en el Programme for International Student Assessment (PISA) 2012, 910 centros educativos participantes.

Lo más relevante de este ejercicio es que es una evaluación diagnóstica, por lo que el uso pretendido es la comprobación del grado de dominio de las destrezas, capacidades y habilidades en expresión y comprensión oral y escrita, cálculo y resolución de problemas de acuerdo con el grado

de adquisición de la competencia en comunicación lingüística y de la competencia matemática de los alumnos que se evalúan. De este modo, la evaluación se centra en los siguientes usos y usuarios:

- Completar y enriquecer la información y la orientación que reciben los alumnos, sus familias y los centros educativos.
- Facilitar que todos los alumnos reciban las medidas específicas de apoyo que sean necesarias.
- Proporcionar información a los centros educativos, para la revisión de los procesos de enseñanza y aprendizaje, y la mejora educativa a partir de los resultados.
- Permitir mejorar a través de la disposición de más información y de la retroalimentación entre evaluación y práctica docente, con el objetivo de tomar las decisiones más acertadas.

Tras la reflexión realizada en el centro educativo sobre las fortalezas y debilidades percibidas en el análisis y la valoración de los resultados de la evaluación, así como la identificación de posibles actuaciones y propuestas de mejora, es necesario seleccionar y priorizar las medidas que los equipos docentes puedan proponer, tanto en el ámbito de desarrollo curricular y metodológico como en el ámbito organizativo y de funcionamiento.

La responsabilidad de la elaboración del plan de mejora corresponde al equipo directivo del centro educativo, y en su producción deberán colaborar el profesorado del centro y los equipos docentes. El plan de mejora formará parte de la Programación General Anual (PGA) del curso siguiente, y su aprobación es una responsabilidad del claustro de profesores de acuerdo con las competencias que se le asignan en el artículo 129 de la ley.

En cualquier caso y de manera general, estos informes del centro educativo se elaboran con carácter diagnóstico y no fiscalizador. Deben servir para detectar carencias y elaborar planes de mejora. No se usan para rendir cuentas, para la ordenación o la selección de personal, ni para dar incentivos o ejecutar sanciones.

Allá donde tiene competencia el MECD los resultados por centro educativo, así como sus informes no se hacen públicos, son de uso interno. En el ámbito competencial de las comunidades autónomas cada una toma la decisión que cree más oportuna para su gestión educativa.

### **3.2.2. Herramientas para apoyar el uso de las evaluaciones**

De manera general el MECD publica en su página de Internet todos los informes relevantes que tratan sobre evaluaciones a gran escala, aunque la mejor recopilación de esta información se puede encontrar en la página web del INEE (ver: <http://www.mecd.gob.es/inee/portada.html>).

La página web está diseñada dentro del marco de los servicios informáticos del MECD y en ella se pueden encontrar enlaces a toda la información de las evaluaciones en las que participa el INEE, sean nacionales o internacionales. Esta sección de la página web se divide en siete partes principales: informes y publicaciones, bases de datos, herramientas gráficas, reportes escolares, organización de eventos, redes sociales, y programa de visitas. A continuación, se proporciona información más detallada sobre cada una de ellas.

**Bases de datos** (<http://www.mecd.gob.es/inee/Bases-de-datos.html>). En este apartado se encuentran las bases de datos de todos los estudios internacionales en los que España participa o ha participado en algún momento:

- PISA en todos sus ciclos.
- TIMSS en 1995 y 2011.
- PIRLS en 2006 y 2011.
- Encuesta Internacional sobre Docencia y Aprendizaje (TALIS, por sus siglas en inglés) en 2008 y 2013.
- Programa para la Evaluación Internacional de Competencias de Adultos (PIAAC, por sus siglas en inglés) en 2012.
- Estudio Internacional sobre la Formación Inicial en Matemáticas (TEDS-M, por sus siglas en inglés) en 2008.
- Estudio Europeo de Competencia Lingüística.
- Evaluación de programas de acción educativa: Programa de Refuerzo, Orientación y Apoyo (PROA) y Escuela 2.0.
- Evaluación General de Diagnóstico de primaria 2009.
- Evaluación General de Diagnóstico de secundaria 2010.

Estas bases de datos se encuentran libres para que cualquier investigador pueda usarlas y hacer análisis secundarios con ellas. En el momento que se sube alguna nueva base de datos se envía un correo electrónico masivo a todos los investigadores con los que el INEE ha colaborado en algún momento para informar de esa nueva disposición, y se indica el enlace a la base de datos y el contenido de la misma. Si fuera preciso, se suele subir también un manual de uso de éstas.

**Informes y publicaciones.** Como su nombre lo sugiere en este apartado se encuentran direcciones para los informes, publicaciones y boletines de educación en los que el INEE se encuentra involucrado de alguna manera.

- **Informes** ([http://www.mecd.gob.es/inee/Ultimos\\_informes.html](http://www.mecd.gob.es/inee/Ultimos_informes.html)). Cada uno de los últimos informes publicados tiene un vínculo. Si se accede a cualquiera de los enlaces que aparecen se abre a una web específica de cada informe. Por ejemplo, en PIRLS-TIMSS 2011 se encontraron los informes nacionales primario y secundario, este último traducido al inglés, además de un boletín informativo de cuatro hojas, así como los informes internacionales. Se tiene acceso a una aplicación interactiva con las preguntas liberadas de TIMSS y PIRLS (<http://evaluacion.educalab.es/timsspirls/>).
- **Publicaciones** (<http://www.mecd.gob.es/inee/publicaciones.html>). Se publican todos los informes, marcos teóricos, preguntas liberadas, etcétera, de todas las evaluaciones nacionales e internacionales en las que participa España, así como los indicadores educativos.
- **Boletines de educación.** En esta sección se pueden encontrar boletines de educación de cinco páginas diferentes. Estos documentos, muy gráficos y directos, son producidos por el INEE, o bien por otros organismos internacionales. En caso de ser necesario estos documentos son traducidos por el INEE.
  - EducaINEE (<http://www.mecd.gob.es/inee/Boletin-de-educacion.html>).
  - PISA in Focus (<http://www.mecd.gob.es/inee/PISA-in-focus.html>).

- Education Indicators in Focus (<http://www.mecd.gob.es/inee/Indicators-in-focus.html>).
- Teaching in Focus (<http://www.mecd.gob.es/inee/Teaching-in-Focus.html>).
- Policy Brief Series (<http://www.mecd.gob.es/inee/Policy-Brief-Series.html>).

Todos estos informes y publicaciones van dirigidos a la comunidad educativa en general: familias, educadores, administración educativa, investigadores, etcétera. Son informes y publicaciones rigurosos, pero de fácil lectura y comprensión para que puedan llegar a todos los actores del sistema educativo.

**Herramientas gráficas** (<http://www.mecd.gob.es/inee/Preguntas-liberadas.html#Otras>). Este tipo de herramientas se han generado para proveer al docente de recursos que le sean útiles cotidianamente. Se han puesto a disposición los ítems liberados de los distintos estudios internacionales tratados con anterioridad mediante plataformas interactivas. No existe uniformidad de las mismas, pues se han ido desarrollando en distintas fases y por distintos agentes. Algunos de estos ítems liberados también se encuentran en documentos planos en formato PDF.

En general se trata de herramientas muy ilustrativas que permiten al docente trabajar con pruebas interactivas que evalúan competencias. En esta plataforma los docentes pueden usar los ítems liberados de PISA, PIRLS, TIMSS, PIAAC y el Estudio Europeo de Competencia Lingüística (EECL). Es muy recomendable acceder a estas herramientas.

Cada vez que se actualiza alguna de estas herramientas se envía un correo electrónico masivo a todos los centros educativos para que conozcan las novedades. Además, en todos los congresos, ponencias, simposios, talleres, etcétera, organizados por el INEE, se hace difusión de éstas.

**Reportes escolares.** Aunque ya se ha hablado de ellos con anterioridad, es importante recordar que para las evaluaciones censales se genera un informe por alumno, un informe por centro educativo y otro para la administración competente en política educativa. Estos informes son elaborados por el INEE y enviados a diversos destinatarios (alumnos, familias, profesores, centros educativos y hacedores de políticas educativas). Son muy ilustrativos y fácilmente entendibles; no obstante, es costumbre del INEE enviar a técnicos para que expliquen en ponencias monográficas el modo de interpretarlos. Con ello se consigue que se usen para generar planes de mejora de la calidad del sistema educativo de modo global.

**Organización de eventos.** Los eventos organizados por el INEE en esta área son de cinco tipos principalmente: presentación de informes, ponencias y comunicaciones, congresos y simposios, talleres y MOOC.<sup>1</sup>

- **Presentaciones oficiales de informes.** Cada vez que termina un ciclo de una evaluación nacional o internacional se hacen presentaciones oficiales por parte de los hacedores de políticas educativas del más alto nivel: ministro(a) de Educación, secretario(a) de Educación o director(a) General de Evaluación. Entre las últimas presentaciones, cabe destacar:

<sup>1</sup> Los MOOC (siglas en inglés de *Massive Open Online Course*) son cursos en línea dirigidos a un amplio número de participantes a través de Internet según el principio de educación abierta y masiva.

- Informes de Evaluaciones Generales de Diagnóstico primaria y secundaria (2009-2010).
- Informes PISA.
- Informes TIMSS-PIRLS.
- Informe TALIS.

Normalmente se dirige la rueda de prensa a medios de comunicación de ámbito nacional, y se presenta toda la documentación relacionada con el estudio en cuestión. La repercusión siempre ha sido alta debido a la importancia del tema por presentar y al valor que la autoridad competente proporciona.

Por tanto, esta información no se dirige exclusivamente a la comunidad educativa, sino que la población destino es todo el Estado. Este altavoz se necesita para hacer hincapié en la importancia de las evaluaciones dentro del sistema educativo.

- **Ponencias y comunicaciones** (<http://www.mecd.gob.es/inee/Ponencias.html>). Durante los últimos años se ha llevado a cabo un ciclo de ponencias y comunicaciones de investigadores mundiales de primer nivel. Por este ciclo han pasado personalidades como Steve Rivkin, Irwin Kirsch, Pablo Zoido o Eric Hanushek, entre otros. El aforo para estas ponencias es muy limitado, pero se retransmiten vía *streaming* y se deja colgado el video a disposición de la comunidad educativa durante varios meses. Además, se envía un correo electrónico masivo a todos los investigadores y actores educativos para invitarlos como asistentes o como público de la transmisión por Internet. El objetivo final es promover el uso de las bases de datos entre los investigadores para continuar trabajos relacionados con las evaluaciones educativas a gran escala.
- **Congresos y simposios.** Igualmente, es habitual organizar congresos posteriores a la presentación oficial de los informes. En estos congresos se invita a docentes, equipos directivos, investigadores y administradores públicos. Como medida adicional se comunica mediante correos electrónicos masivos a cualquier persona u organismo que pudiera estar interesado en el congreso, y éste se retransmite vía *streaming* para tener más alcance. La estructura de estos congresos siempre es similar: presentación de informes y datos por técnicos del INEE, análisis secundarios por investigadores invitados, mesas redondas compuestas por administradores de políticas educativas, o por centros con buenas prácticas y, como eje central, la participación de los asistentes en debates sobre los temas tratados. Quizá, entre los últimos congresos, los más relevantes han sido los siguientes:
  - Congreso PIRLS-TIMSS: presentación de informes de centro.
  - Congreso TALIS.
  - PISA: Evaluación por Ordenador y Resolución de Problemas, Matemáticas y Lectura.
  - PISA. Finanzas para la vida.
  - El último simposio organizado ha sido: Ciencias e inglés en la evaluación internacional.
- **Talleres.** El INEE promueve dos modelos de talleres. Por un lado, aquellos dirigidos a técnicos de las administraciones educativas con el fin de formarlos en el análisis de datos de evaluaciones a gran escala, la elaboración de informes y análisis secundarios. Y, por otro, talleres dirigidos a docentes en general en los cuales se les forma en la elaboración de pruebas de evaluación a gran escala en competencias. Del primer tipo de talleres cabe destacar tres:

- Uso de la Teoría de la Respuesta al Ítem (TRI) en evaluaciones a gran escala.
- Análisis de datos de las evaluaciones internacionales a gran escala PISA, TIMSS, PIRLS, etcétera.
- Análisis multinivel en evaluaciones.

Del segundo tipo de taller se destacan:

- Presentación de ítems liberados de PISA, TIMSS y PIRLS: marco teórico, elaboración y aprovechamiento didáctico.
- Ciencias e inglés en la evaluación internacional.

El INEE tiene la consigna de que no hay mejor manera de promocionar las evaluaciones del sistema educativo a gran escala que inculcar en el profesorado la cultura de la evaluación; mostrando mediante talleres las bondades de este tipo de estudios, se consigue que el docente entienda las evaluaciones como un apoyo y no como un enemigo. Esto lleva a que los talleres sean impartidos por expertos internacionales en cada una de las competencias junto con técnicos del INEE con amplia experiencia en evaluación.

- **MOOC.** Las preguntas de PISA y TIMSS: una práctica para el aula. Nada hay más abierto y popular que un MOOC para que toda la comunidad educativa conozca las evaluaciones a gran escala. Este MOOC en particular estuvo dirigido al público interesado en las evaluaciones a gran escala y tuvo más de diez mil inscripciones (de las cuales cerca de la mitad correspondió a maestros de América Latina). Se hizo un repaso completo de lo que suponen las evaluaciones internacionales; los puntos tratados fueron:
  - Finalidad de las pruebas externas.
  - Las pruebas externas en los medios.
  - Los ítems liberados.
  - Corrección de pruebas externas.
  - De la pregunta al informe: datos.
- **Redes sociales.** El INEE se encuentra representado en las redes sociales más populares; tiene la consigna de que no estar en la red es no estar; por tanto, tiene con las siguientes cuentas:
  - **Twitter:** @educalNEE
  - **YouTube:** El INEE posee su propio canal en YouTube (<https://www.youtube.com/channel/UCBKvUbEeC56etpkHVLPbxhA/videos>); en este canal hay una gran colección de videos explicativos de distintos ámbitos de las evaluaciones a gran escala; los actores de estos videos son técnicos del INEE, investigadores, administradores educativos, docentes, etcétera. Todos han sido grabados bajo la supervisión del INEE y puestos a disposición los interesados en el tema.
  - **Slideshare:** Igual que en YouTube, el INEE posee canal propio en Slideshare (ver: [http://es.slideshare.net/INEE\\_MECD](http://es.slideshare.net/INEE_MECD)); a este canal se suben las presentaciones que se han usado en los congresos, talleres, ponencias, etcétera, en las que participan los técnicos del INEE o cualquier otro colaborador que dé su permiso expreso. Este canal está a disposición de cualquier persona que tenga interés en las evaluaciones.

- **Programa de visitas.** Para terminar con las herramientas que el INEE pone a disposición del público se debe citar el programa de visitas. El INEE invita a investigadores, administradores de la educación y profesores a realizar estancias en su sede para intercambiar información relacionada con la evaluación educativa. El perfil de los visitantes es de personas interesadas en la evaluación educativa y que estén avaladas por un organismo o una institución de reconocido prestigio. Excepcionalmente, también se admite a aquellas personas que, de manera individual, presenten un *curriculum vitae* que garantice una trayectoria relacionada con la evaluación educativa.

Los objetivos fundamentales de este programa son dos: por un lado, fomentar la colaboración entre las personas interesadas en las investigaciones y los estudios de evaluación del sistema educativo y las personas que trabajan en el INEE. Y por otro, crear lazos de colaboración y fortalecer la cooperación nacional e internacional con organismos e instituciones de prestigio en el mundo educativo. Los visitantes tendrán la posibilidad de realizar jornadas informativas para exponer al personal del INEE las investigaciones en las que están trabajando o las actuaciones que realizan en su centro u organismo de procedencia.

### 3.2.3. Uso de los resultados

Como se ha señalado, esta evaluación es diagnóstica, con datos de uso interno: sus bases de datos no se hacen públicas. Como consecuencia, no se pueden recoger estadísticas adicionales ni realizar análisis secundarios de ningún tipo. Son muy pocas comunidades autónomas las que hacen públicas algunas estadísticas y generan un *ranking* de centros en función de los resultados obtenidos en estas pruebas. Como ejemplo más significativo se destaca la comunidad autónoma de Madrid, que lo publica en la siguiente página web ([www.mecd.gob.es/inee/Evaluacion\\_sexto\\_Primeria.html](http://www.mecd.gob.es/inee/Evaluacion_sexto_Primeria.html)).

Si se focaliza en el uso interno de los resultados por centro educativo y por alumno estos llevan un informe orientador junto con el boletín de calificaciones anuales, más el informe proporcionado por el MECD. Los centros educativos incorporan en su PGA un plan de mejora una vez visto el informe proporcionado. Esta PGA se comunica a toda la comunidad educativa del centro en cuestión.

### 3.2.4. Opinión pública

La realimentación que recibe el INEE de la opinión pública en general se restringe a los medios ya mencionados: presentación de ruedas de prensa, y la administración de redes sociales y de canales propios en diversas plataformas en Internet.

Otro medio importante para obtener realimentación sobre el uso de sus herramientas es el blog propio del INEE (<http://blog.educalab.es/inee/>), en el que periódicamente se suben artículos relacionados con situaciones de actualidad en torno a la evaluación.

Por este medio el INEE ha recabado información que sugiere que la opinión pública no entiende bien el significado de las evaluaciones a gran escala, y que la interpretación de los datos no suele ser correcta. En el INEE existe un compromiso por intentar explicar técnicamente qué son, qué miden, y para qué valen las evaluaciones, pero la evidencia sugiere que el objetivo no está del

todo cumplido, al menos en lo que se refiere al público no especializado y a la opinión pública en términos más generales.

De acuerdo con la retroalimentación recibida por este organismo, se tiene la certeza de que un resultado positivo es que los docentes en España perciben las evaluaciones como una ayuda importante para su trabajo.

### 3.2.5. Ejemplos de materiales de apoyo

A continuación, se presenta una serie de direcciones para diferentes páginas web y recursos en línea que utiliza o difunde el INEE en relación con el uso de las evaluaciones de gran escala en España.

- Portal web INEE: <http://www.mecd.gob.es/inee/portada.html>
- Informe Evaluación General de Diagnóstico de educación primaria 2009: <http://www.mecd.gob.es/dctm/ievaluacion/evaluaciongeneraldiagnostico/pdf-completo-informe-egd-2009.pdf?documentId=0901e72b8015e34e>
- Informe Evaluación General de Diagnóstico de educación secundaria 2010: <http://www.mecd.gob.es/dctm/ievaluacion/informe-egd-2010.pdf?documentId=0901e72b80d5ad3e>
- Página web de la evaluación de tercer curso de educación primaria. En ella se puede encontrar el marco general de la evaluación, un video explicativo, entradas al blog, boletines, modelos de pruebas y pruebas pasadas definitivamente, además de otra información relevante: [http://www.mecd.gob.es/inee/Evaluacion\\_tercero\\_Primeria.html](http://www.mecd.gob.es/inee/Evaluacion_tercero_Primeria.html)
- Página web de la evaluación de final de educación primaria. En ella se puede encontrar el marco general de la evaluación, una presentación en Power Point explicativa, entradas al blog, boletines y modelos pruebas, además de otra información relevante: [http://www.mecd.gob.es/inee/Evaluacion\\_sexto\\_Primeria.html](http://www.mecd.gob.es/inee/Evaluacion_sexto_Primeria.html)
- Enlace a bases de datos de los estudios nacionales e internacionales en los que participa el INEE: <http://www.mecd.gob.es/inee/Bases-de-datos.html>
- Enlace a últimos informes publicados: [http://www.mecd.gob.es/inee/Ultimos\\_informes.html](http://www.mecd.gob.es/inee/Ultimos_informes.html)
- Preguntas interactivas de TIMSS-PIRLS: <http://evaluacion.educalab.es/timsspirls/>
- Publicaciones: <http://www.mecd.gob.es/inee/publicaciones.html>
- Enlaces a los boletines de EducaINEE, PISA in Focus, Indicators in Focus, Teaching in Focus y Policy Brief Series:
  - <http://www.mecd.gob.es/inee/Boletin-de-educacion.html>
  - <http://www.mecd.gob.es/inee/PISA-in-focus.html>
  - <http://www.mecd.gob.es/inee/Indicators-in-focus.html>
  - <http://www.mecd.gob.es/inee/Teaching-in-Focus.html>
  - <http://www.mecd.gob.es/inee/Policy-Brief-Series.html>
- Preguntas liberadas de los estudios internacionales; en este enlace aparecen preguntas en PDF y en formato interactivo: <http://www.mecd.gob.es/inee/Preguntas-liberadas.html#Otras>
- Información de las ponencias y comunicaciones que se han llevado a cabo en el INEE hasta el momento: <http://www.mecd.gob.es/inee/Ponencias.html>
- Twitter: @educalNEE
- Canal YouTube: <https://www.youtube.com/channel/UCBKvUbEeC56etpkHVLPbxhA/videos>

- Canal Slideshare: [http://es.slideshare.net/INEE\\_MECD](http://es.slideshare.net/INEE_MECD)
- Blog del INEE: <http://blog.educalab.es/inee/>
- Ejemplos de informes, en documentos adjuntos:
  - Modelo informe centro
  - Modelo informe alumno\_Cmat
  - Modelo informe alumno\_Cling
- Folletos para congresos, talleres, etcétera:
  - Simposio Ciencias e Inglés en la Evaluación Internacional: <http://www.mecd.gob.es/dctm/inee/simposio---ciencias-e-ingles/triptico-simposio.pdf?documentId=0901e72b81ef22c9>
  - Congreso PIRLS-TIMSS: presentación de Informes de centro educativo: <http://www.mecd.gob.es/dctm/inee/internacional/congresopirls-timss-2013/programa-pirls-timss.pdf?documentId=0901e72b8175e552>
  - Congreso TALIS: <http://www.mecd.gob.es/dctm/inee/internacional/congresotalis2013/programa-talis-definitivo.pdf?documentId=0901e72b819e1724>
  - Congreso PISA Finanzas para la Vida: <http://www.mecd.gob.es/dctm/inee/congresos/pisa-finanzasparalavida/programa-140708.-congreso-pisafinanzasparalavida.pdf?documentId=0901e72b81a28e74>
  - Congreso PISA Evaluación por Ordenador y Resolución de Problemas: <http://www.mecd.gob.es/dctm/inee/congresos/pisa-2012-jornadas-abril-2014/diptico-congreso-pisa-tit-modif-01-4-2014.pdf?documentId=0901e72b8190479f>
  - Seminario Presentación de Ítems Liberados de PISA, TIMSS y PIRLS: Marco teórico, elaboración y aprovechamiento didáctico: <http://www.mecd.gob.es/dctm/inee/cursos/comillas-programaactualizado9agosto13.pdf?documentId=0901e72b817145a>

# 4

## El caso de los Estados Unidos de América

---

### 4.1. Características de la evaluación de gran escala

#### 4.1.1 Introducción

La Evaluación Nacional de Logro Educativo (NAEP, por sus siglas en inglés) es un proyecto de evaluación designado por el Congreso de los Estados Unidos en el que se evalúan los resultados de aprendizajes de los estudiantes en diversas asignaturas. Es la evaluación “con mayor representatividad nacional y continua” (NCES, 2013) de los aprendizajes de los estudiantes de ese país.

Esta evaluación se comenzó a aplicar a partir de 1969, aunque la preparación para dicha aplicación comenzó cinco años antes (Ha, 2006: 1). Fue propuesta en 1966 por Francis Keppel y Ralph Tyler, quienes plantearon la necesidad de una evaluación nacional que “diera información acerca del rendimiento académico de los alumnos en la nación” (Ha, 2006: 17).

La periodicidad de la aplicación de las pruebas ha variado a través del tiempo. Durante los primeros diez años (1969-1979) se administraba al menos una prueba por año; entre 1980 y 1996 las evaluaciones se realizaron cada dos años, y a partir de 1996 las aplicaciones se volvieron a realizar anualmente (NCES, 2009: 3).

El propósito principal de NAEP es medir el logro educativo de los estudiantes de los Estados Unidos (NCES, 2009: 1). Se busca también monitorear el progreso y los cambios en el logro educativo a través del tiempo; por esta razón las pruebas NAEP han sido diseñadas para garantizar que sean comparables entre años (IES, 2015).

La documentación de NAEP explicita que los reportes buscan “describir el desempeño de los estudiantes en el contexto del sistema educativo para informar las discusiones de tomadores de decisiones y líderes educativos” (NCES, 2009: 1). No busca “promover políticas o pedagogías específicas, o dirigir el establecimiento de estándares educativos, pruebas y currículos estatales y locales” (NCES, 2009: 1).

Originalmente NAEP tenía como objetivo proporcionar información sobre el desempeño de los alumnos a nivel nacional; sin embargo, en 1990 el Congreso autorizó que las evaluaciones de

NAEP se realizarán a nivel estatal para poder realizar comparaciones entre estados y con los resultados a nivel nacional (NCES, 2009: 3).

NAEP tiene dos grandes componentes: NAEP principal y NAEP de tendencias a largo plazo. El primer componente incluye evaluaciones con representatividad estadística nacional, estatal y de algunos distritos urbanos “que utilizan los mismos instrumentos, basados en los marcos de referencia más recientes” (NCES, 2009: 4). Las pruebas nacionales se aplican a estudiantes de cuarto, octavo y doceavo grados; las estatales, a estudiantes de grados cuarto y octavo, y las de distritos urbanos, a alumnos de cuarto y octavo grados que se encuentran estudiando en escuelas públicas.

El componente de tendencias a largo plazo de NAEP administra pruebas sólo a nivel nacional utilizando instrumentos que permiten la comparación a través del tiempo. Las pruebas se aplican a estudiantes de 9, 13 y 17 años.

Los resultados de NAEP se publican en la Libreta de Calificaciones de la Nación (The Nation’s Report Card), y proporcionan resultados nacionales, de estados y de distritos urbanos seleccionados (IES, 2016).

En la documentación oficial de NAEP se identifica como usuarios de la información a investigadores, diseñadores de políticas y educadores, entre otros. Estos usuarios pueden estar presentes a niveles nacional, estatal y distrital, y en jurisdicciones que son representadas en las diferentes pruebas. Se afirma que la información que genera NAEP puede ser sujeta a múltiples análisis secundarios (NCES, 2009: 33). En particular se precisa que los resultados pueden ser utilizados para la mejora de la educación de los Estados Unidos por el siguiente grupo de actores: “legisladores, educadores a nivel estatal y local, directores, maestros y padres” (NCES, 2010c).

Por otro lado, desde 2001 la legislación No Child Left Behind (NCLB) estableció como requisito que los estados participen en NAEP estatal (grados cuarto y octavo; lectura y matemáticas; cada dos años). La participación de los estados en esta prueba se considera necesaria para obtener información que permita valorar el progreso de los estudiantes a través del tiempo. Este requisito hace que NAEP se vea también como un sistema de evaluación ligado a la rendición de cuentas; si los estados no participan, no son elegibles para recibir financiamiento bajo el Acta de Educación Primaria y Secundaria (ESEA, por sus siglas en inglés) (Ha, 2006: 3).

#### **4.1.2. Marco administrativo/legal**

La prueba es administrada por el Centro Nacional para Estadísticas de la Educación (NCES, por sus siglas en inglés). El NCES se encuentra dentro del Instituto de Ciencias de la Educación del Departamento de Educación de los Estados Unidos. La operación del NCES depende de las políticas emitidas por la Junta Regidora de la Evaluación Nacional (NAGB, por sus siglas en inglés). El Comisionado Asociado forma parte del NCES, y es quien está a cargo de ejecutar el programa de operaciones y asegurar el control de calidad de la prueba.

La NAGB es una organización independiente, creada por el Congreso en 1988, para supervisar y regular la operación de NAEP a través del establecimiento de políticas. La NAGB está integrada por 26 personas, quienes se reúnen trimestralmente para tomar decisiones sobre NAEP, dentro de las que se incluyen: determinar las materias, el contenido y los niveles de logro a evaluar, así

como los marcos de referencia de las evaluaciones; aprobar todas las preguntas que los estudiantes responderán; establecer el cronograma de las evaluaciones; comunicar resultados a diferentes audiencias, y tomar medidas para mejorar el reporte de resultados (NAGB, s. f.).

Las funciones del NCES son: “desarrollar las preguntas de la prueba, evaluar a los estudiantes, calificar las pruebas, analizar los datos obtenidos y presentar los resultados” (NCES, 2010c: 1).

Los estados participan de forma voluntaria en NAEP; sólo en las pruebas de lectura y matemáticas, que se aplican a cuarto y octavo grados cada dos años, la participación es obligatoria debido al Acta de No Child Left Behind (NAGB, s. f.). Sin embargo, la participación de los estudiantes en la NAEP es voluntaria (NAGB, s. f.).

### 4.1.3. Procedimientos y metodología

En el desarrollo de la evaluación participan múltiples organizaciones (NCES, 2015b), y ciertamente, la participación de las escuelas, los distritos escolares y personal de NAEP es fundamental para que la prueba pueda ser aplicada.

El diseño, el análisis y el reporte de resultados de las pruebas está a cargo del Educational Testing Service (ETS) desde 1983, con base en los marcos de referencia diseñados por la NAGB.

NAEP realiza pruebas muestrales. Los diseños muestrales varían de acuerdo al alcance de las evaluaciones, es decir, si tienen representatividad estadística estatal o nacional. En ambos casos las muestras son probabilísticas (cada escuela y alumno tiene probabilidad de ser seleccionado), y se incluyen alumnos de escuelas públicas y privadas. A continuación, se describen las características generales de los diferentes tipos de diseños muestrales:

**Diseño muestral y selección de escuelas públicas en las aplicaciones estatales.** El diseño muestral es de etapas múltiples: selección de escuelas públicas dentro de áreas designadas; selección de estudiantes en los grados a evaluar dentro de las escuelas, y repartición de los estudiantes en las asignaturas por evaluar (NCES, 2011). Para la selección de las escuelas se lleva a cabo un proceso de estratificación dentro de cada estado tomando en cuenta sus características. La selección de escuelas considera diferentes estratos. Los estratos están vinculados a diferentes características de las escuelas que se relacionan con el logro educativo, por ejemplo: “su ubicación, la cantidad de alumnos inscritos que pertenecen a grupos minoritarios, puntajes de logro estatales, la mediana de los ingresos del área en que se ubica la escuela” (NCES, 2011).

**Diseño muestral y selección de escuelas privadas en evaluaciones estatales.** La muestra de escuelas privadas no se diseña para generar estimadores a nivel estatal; sólo se pueden hacer inferencias a nivel nacional. Se hace una clasificación de las escuelas privadas por tipo, y se agrupan tomando en cuenta su ubicación geográfica, su grado de urbanidad y la cantidad de alumnos inscritos (NCES, 2011).

**Diseño muestral en las evaluaciones nacionales.** El diseño también tiene múltiples etapas. La primera etapa consiste en la selección de unidades geográficas muestrales primarias (*geographic primary sampling units*; aproximadamente se conforman mil unidades). Después de la selección de estas unidades, se seleccionan muestras de escuelas públicas y privadas siguiendo procedimientos de elección similares a los descritos en los diseños estatales (NCES, 2011).

El levantamiento de datos se hace durante el horario escolar habitual. La aplicación tiene una duración de entre 90 y 120 minutos. Se aplican las pruebas de logro educativo y cuestionarios de contexto. Los cuestionarios de contexto los responden alumnos, maestros y el director de la escuela.

El levantamiento de la información es un proceso que requiere la colaboración de "las escuelas, distritos escolares, estados y el personal de NAEP" (NCES, 2014b). El personal de campo de NAEP es responsable de realizar la mayoría de las acciones para la administración de la prueba (NCES, 2015a). Se encarga de entregar y recolectar los materiales directamente en las escuelas. Emplea procedimientos que buscan garantizar la integridad de los resultados de las evaluaciones y reducir la carga que representa la participación en estos procesos de evaluación para las escuelas (NCES, 2015a).

El personal de campo comprende: coordinadores estatales de NAEP, directores de campo, gerentes de campo, supervisores, coordinadores de la evaluación, y administradores de la evaluación (NCES, 2014a). Quienes participan directamente en la administración de las pruebas son los coordinadores y administradores de la evaluación. Los primeros se encargan de establecer contacto con las escuelas antes de la aplicación; distribuir y recolectar los materiales; registrar las decisiones sobre los acomodos a implementar y planear con ese fin; programar las actividades de evaluación de acuerdo con las necesidades de las escuelas; identificar el personal necesario para la aplicación conforme al horario acordado; dar seguimiento a la capacitación de los aplicadores (administradores de la prueba); coordinar las actividades durante la aplicación, e implementar controles de calidad en cuanto a los materiales. Los aplicadores o administradores de las evaluaciones desarrollan tareas administrativas (por ejemplo, la preparación de cuadernillos y materiales); revisan la asistencia; leen indicaciones a los alumnos; aplican las pruebas, y garantizan la precisión, la integridad y la confidencialidad de datos y materiales (NCES, 2014a).

El diseño muestral y la recolección de datos está a cargo de Westat desde 1983, comprende: "el muestreo de escuelas y alumnos; reclutamiento de escuelas; capacitación a personal de campo; aseguramiento de calidad y elaboración de pesos muestrales" (NCES, 2015b). Westat también está a cargo del centro de apoyo y servicio desde 2002, dedicado a ofrecer apoyo y capacitación a los coordinadores estatales de NAEP y la evaluación de distritos urbanos (NCES, 2015b).

La población objetivo de NAEP varía de acuerdo al tipo y el alcance de la prueba. Las evaluaciones nacionales se aplican a estudiantes de escuelas públicas de cuarto, octavo y doceavo grados, sobre las siguientes asignaturas: "matemáticas, ciencias naturales, escritura, historia de los Estados Unidos, economía, civismo, geografía y las artes" (NCES, 2009: 4), entre otras. Las evaluaciones estatales y a distritos urbanos se aplican en cuarto y octavo grados, en relación con Lectura, Matemáticas, Ciencias y Escritura (NCES, 2009: 4). Las evaluaciones de tendencias a largo plazo se aplican a niños de 9, 13 y 17 años en relación con Matemáticas y Lectura (NCES, 2009: 4).

## 4.2. Características de las prácticas de uso de los resultados de las evaluaciones

### 4.2.1 Introducción

En la documentación de NAEP se señala que los resultados de las pruebas pueden ser utilizados por múltiples actores para "evaluar el progreso de los estudiantes, y desarrollar nuevas estrategias para

mejorar la educación en los Estados Unidos” (NCES, 2010a: 3). Dentro de los diferentes actores educativos a los que se hace alusión se encuentran: investigadores, diseñadores de políticas, tomadores de decisiones, educadores (maestros y directores) y padres de familia. Los usos posibles de los resultados se vinculan con los dos grandes objetivos de NAEP: “comparar el logro de los estudiantes entre estados y otras jurisdicciones y monitorear el cambio en el logro entre estudiantes de cuarto, octavo y doceavo grados” (NCES, 2012).

En la página de NAEP se presenta información para los siguientes tipos de usuarios: escuelas seleccionadas, padres de familia, estudiantes, educadores investigadores, medios de comunicación y (NCES, 2010b). A los cuatro primeros se les muestra información que permita tener mayor claridad sobre las pruebas y lo que representa la participación de las escuelas y los estudiantes en ellas (por ejemplo, se precisan tiempos de aplicación, se muestran diferentes tipos de preguntas, etcétera); también se proporcionan alternativas para explorar los resultados. A los investigadores se les describen las bases de datos y herramientas de análisis disponibles, y se les muestran diferentes recursos a los que pueden tener acceso (por ejemplo, documentación técnica, capacitación, biblioteca). A los medios de comunicación se les facilita el acceso a los comunicados de prensa, tablas con resultados, videos, capacitaciones o encuentros en línea, entre otros recursos.

No se especifican los usos esperados para los diferentes tipos de audiencias; sólo en algunos casos se encontraron descripciones sobre los usos que podrían hacerse. Por ejemplo, se especifica que los educadores pueden utilizar los marcos de referencia para diseñar evaluaciones y revisar el currículo; también se sugiere que las preguntas de respuesta construida y las guías para la calificación podrían servir como modelo para desarrollar nuevas prácticas de evaluación en aula (NCES, 2012).

En la evaluación que se hizo a NAEP en 2009 (Buckendahl *et al.*, 2009a) se reconocen tres principales usos de los resultados: a) monitorear las tendencias en el logro de los estudiantes; b) evaluar el nivel de desempeño de los alumnos, y c) hacer comparaciones de los resultados entre estados. A estos usos se añade el reporte de resultados por distritos, que es el más reciente. Estos usos se podrían clasificar dentro de las siguientes categorías: descriptivos (primer uso), evaluativos (segundo uso) y comparativos (tercer y cuarto usos). El primero, la descripción del logro educativo, ha sido el más sostenido a través del tiempo, para esto ha requerido que se mantengan características del diseño de las pruebas. La evaluación del nivel de desempeño de los alumnos ha sido principalmente cuestionada por la exigencia de los niveles establecidos en las pruebas. Los usos comparativos también han sido cuestionados: las comparaciones entre estados parecen tener como limitación la subrepresentación de algunos grupos poblacionales; las comparaciones entre distritos aún requieren evidencias de validez (Buckendahl *et al.*, 2009a: XIV, XV).

En la evaluación de NAEP de 2009 se mencionan usos emergentes vinculados con la rendición de cuentas en los estados y con las comparaciones con resultados de evaluaciones internacionales. Con respecto a la rendición de cuentas de los estados, se ha identificado que, a pesar de las diferencias entre las pruebas estatales y NAEP (por ejemplo, estándares de contenidos, características de población, procedimientos, entre otros), se ha tratado de utilizar los resultados de la NAEP para verificar los de las pruebas estatales, lo cual no es posible. En cuanto a las comparaciones con pruebas internacionales, como TIMSS o PISA, en la NAEP se hace un llamado para saber si dichas evaluaciones pueden ser una fuente de validez externa.

#### 4.2.2. Herramientas para apoyar el uso de las evaluaciones

Existen siete herramientas principales para apoyar el uso de los resultados de las evaluaciones de NAEP: el explorador de datos NAEP, ejemplos de preguntas, comparaciones estatales, perfiles estatales, perfiles de los distritos, bases de datos para análisis secundarios, y recursos en otras páginas web. La fuente principal para la identificación de estas herramientas fue la página web de NAEP (<http://nces.ed.gov/nationsreportcard/>). A continuación, se describe cada una de las herramientas encontradas.

##### Explorador de datos NAEP (*NAEP Data Explorer, NDE*)

<http://nces.ed.gov/nationsreportcard/naepdata/>

Es una herramienta en línea que permite explorar los resultados de las evaluaciones por materias, grados o jurisdicciones. Permite el acceso a cuatro tipos de bases de datos: NAEP principal; tendencias a largo plazo; registro de estudiantes de doceavo grado, materias que cursan y calificaciones; y el Estudio Nacional de Educación Indígena.

Con esta herramienta se pueden crear tablas, gráficas y mapas personalizados de acuerdo con los datos que se seleccionen. También se puede analizar la relación que guardan los resultados con variables del contexto de los alumnos, docentes y escuelas.

No se especifica quiénes participaron en el diseño. De manera general se menciona que esta herramienta puede usarse para analizar los datos de múltiples formas.

Su uso se promueve con algunos recursos que explican su función, el contenido y la forma en que puede ser útil: videos tutoriales, guías de referencia y páginas de ayuda. La guía de referencia describe: el acceso a la herramienta, su uso, ejemplos de usos y los pasos detallados a seguir en cada caso (<http://nces.ed.gov/nationsreportcard/pdf/NDEQuickReferenceGuideWebVersion05.18.09.pdf>). La página de ayuda permite hacer una búsqueda utilizando palabras clave o de referencia de acuerdo con el interés del usuario ([http://nces.ed.gov/nationsreportcard/NDEHelp/WebHelp/NDE\\_Help\\_Main.htm](http://nces.ed.gov/nationsreportcard/NDEHelp/WebHelp/NDE_Help_Main.htm)).

##### Ejemplos de preguntas (*sample questions*)

<http://nces.ed.gov/nationsreportcard/itmrlsx/>

Contiene más de tres mil preguntas que han sido liberadas después de las aplicaciones de NAEP. Se divide a su vez en cuatro tipos de herramientas relacionadas con las preguntas liberadas que, de acuerdo con la información incluida en la descripción de la herramienta, podrían ser utilizados para complementar la enseñanza de los estudiantes, lograr una mayor comprensión de las evaluaciones y su contenido, y conocer los resultados de aprendizaje de los estudiantes. Las herramientas en que se divide son:

- **Questions Tool:** esta herramienta permite explorar las preguntas liberadas por NAEP. La exploración se hace de acuerdo con la asignatura evaluada, el grado en que se aplicó la prueba, el año de aplicación, el área de contenido de la prueba, el dominio cognitivo, el nivel

de dificultad de la pregunta y el tipo de respuesta (por ejemplo, opción múltiple, respuesta construida corta y respuesta construida extendida). Para cada pregunta que resulte de la exploración se muestran la consigna, la guía para calificación, ejemplos de respuestas (particularmente útiles para los reactivos de respuesta construida) y también datos sobre el comportamiento de las respuestas de los alumnos.

Esta herramienta también permite descargar e imprimir las preguntas. Se estima que esta herramienta puede ser particularmente útil para los docentes, pues los criterios de búsqueda permiten identificar preguntas relacionadas con los contenidos que trabajan en sus aulas con diferentes dominios cognitivos y niveles de dificultad. La elaboración de exámenes mediante esta herramienta es muy accesible; los maestros pueden hacer una selección a partir de los resultados de la búsqueda, guardar su selección, y si esta selección conforma completamente los exámenes que se aplicarán, incluso pueden publicarlos en línea o imprimirlos directamente. Cabe señalar que para la publicación y la descarga de las pruebas es necesario que los usuarios se hayan registrado en la plataforma y cuenten con clave de acceso.

Adicionalmente, esta herramienta da la oportunidad de probar las preguntas de NAEP (*Try NAEP Questions*) en la misma página web. Para esto el usuario debe elegir asignatura, grado, el número de preguntas a probar y el nivel de dificultad. A partir de estas decisiones se muestran las preguntas, una a la vez. Al finalizar se presenta el número de aciertos, y los usuarios pueden revisar los resultados de la prueba (se muestran la respuesta elegida y la respuesta correcta) e imprimirlos.

Finalmente, la herramienta contiene una opción específica para elaborar exámenes. Los usuarios pueden seleccionar un examen ya elaborado, o elaborar uno personalizado. Los usuarios registrados en la plataforma pueden generar el listado de los alumnos y el conteo de aciertos, y tener disponibles estas pruebas en diferentes momentos del periodo escolar.

- **Interactive items:** este componente de la herramienta se enfoca en la exploración de preguntas utilizadas en evaluaciones realizadas por computadora o preguntas que requieren trabajo práctico de los alumnos (por ejemplo, trabajar con un *kit* de materiales para la construcción de un circuito eléctrico). Los reactivos se muestran al seleccionar la asignatura y el grado; no hay opción para seleccionar contenido, dominio cognitivo o nivel de dificultad. En los ítems que requieren trabajo práctico de los alumnos se muestra información sobre los criterios de calificación, ejemplos de respuestas de los alumnos y el comportamiento estadístico de las respuestas.
- **Test yourself:** es una herramienta que permite realizar una prueba con preguntas de diferentes asignaturas. A los usuarios se les muestra un mapa de los reactivos de la prueba de acuerdo con el nivel de dificultad y el tipo de preguntas que los alumnos contestaron correctamente. Los usuarios pueden seleccionar dominios de contenido específicos o responder la totalidad de preguntas disponibles de los diferentes dominios de contenido.
- **Item maps:** esta herramienta permite tener acceso a las descripciones de los ítems de acuerdo con los diferentes niveles de la prueba. Se selecciona la asignatura, el año de aplicación y el grado. En el mapa se establecen vínculos de acceso a los ítems; en estas ventanas se presenta la consigna del reactivo, la guía para la calificación, ejemplos de respuestas y el comportamiento estadístico de los resultados. Cuenta con una guía de uso en formato PDF.

### Comparaciones estatales (*State comparisons*)

<http://nces.ed.gov/nationsreportcard/statecomparisons/>

Es una herramienta que proporciona tablas y mapas que permiten hacer comparaciones del puntaje de distintos grupos de alumnos; despliega los resultados por estados y jurisdicciones sobre las que se tiene representatividad estadística. De acuerdo al tipo de búsqueda, se puede presentar información sobre las diferencias entre estados y su significado estadístico. Para la búsqueda se eligen la asignatura, el grado y el año de aplicación. En cuanto al año de aplicación la herramienta especifica si se acomodó de algún modo a alumnos con necesidades educativas especiales. La herramienta da la oportunidad de seleccionar grupos específicos de estudiantes (como por género, grupo étnico, elegibilidad de almuerzo en escuela, percentiles). Los resultados se pueden exportar a Excel. Cuenta con una guía de uso para los usuarios en formato PDF y una descripción general de la herramienta en html.

### Perfiles estatales (*State profiles*)

<http://nces.ed.gov/nationsreportcard/states/>

Esta herramienta muestra datos clave del desempeño de cada estado que participa en NAEP de cuarto y octavo grados, en las asignaturas de matemáticas, lectura, escritura y ciencias naturales; en algunos casos también se incluye información de doceavo grado. La herramienta permite conocer el desempeño de los estudiantes de los estados a través del tiempo; esta información se muestra en gráficas y tablas por asignatura y grado de aplicación de la prueba. Los usuarios deben seleccionar un estado y, posteriormente, asignatura, grado y año de aplicación específica para obtener información que les permita comparar los resultados entre el estado elegido y el resto. La representación visual permite identificar fácilmente si existen diferencias significativas: se muestra un mapa donde se identifica con un color específico el estado elegido; el resto de los estados tiene colores distintos de acuerdo con el tipo de diferencias que hay entre ellos. Además, se presenta información con respecto al promedio nacional, y la descripción por año de aplicación del porcentaje de alumnos que se encontraron en diferentes niveles de desempeño en el estado.

Incluye información demográfica de los alumnos (por ejemplo, alumnos inscritos, registrados en programas con educación individualizada, o de acuerdo al grupo étnico al que pertenecen) y de las escuelas y distritos escolares (como número de escuelas, o tasa de alumnos por docente). Además, brinda información de los reportes realizados por estados en formato PDF.

Para la guía de los usuarios se incluye una página de ayuda donde se describe el proceso para buscar y generar la información deseada y se vincula con otras herramientas de la página de NAEP y del NCES.

### Perfiles de los distritos (*District profiles*)

<http://nces.ed.gov/nationsreportcard/districts/>

Esta herramienta es similar a la de perfiles estatales. Permite comparar los resultados de las pruebas NAEP aplicadas a Distritos Urbanos —NAEP Tribunal Trial Urban District Assessment (TUDA)—. Las representaciones visuales que se muestran son similares a las de los perfiles estatales. Los usuarios deben seleccionar un distrito de interés y a partir de esta selección se muestran

los puntajes promedio de las diferentes asignaturas que han sido evaluadas por grado y año de aplicación. Al seleccionar una aplicación específica, se puede identificar el tipo de diferencias existentes entre distritos. También permite valorar las diferencias con respecto al promedio nacional, grandes ciudades y el estado al que pertenece el distrito en cuestión.

Para la guía de los usuarios se incluye una página de ayuda donde se describe el proceso para buscar y generar la información deseada.

#### Bases de datos para análisis secundarios

Las bases de datos para análisis secundarios no están disponibles en la página web de NAEP. Se requiere que sean solicitadas directamente al NCES, mediante la página web.

Recursos en otras páginas web (*The Nation's Report Card*)  
[www.nationsreportcard.gov](http://www.nationsreportcard.gov)

Se encontró una página web específica con los resultados de la aplicación más reciente de NAEP, en 2015. Se muestran los resultados a nivel nacional, por estados y distritos. También incluye videos con los hallazgos más sobresalientes del estudio, y da la oportunidad para descargar tablas de datos y generar tablas personalizadas con los resultados. Estas tablas incluyen información sobre las tasas de participación, alumnos con alguna discapacidad, y alumnos que se encuentran aprendiendo inglés.

#### 4.2.3. Uso de los resultados

Dentro de las páginas de NCES y la específica de *The Nation's Report Card* no se encontraron estadísticas de uso de la información. Sin embargo, se encontraron cuatro estudios relacionados con la presentación o el uso de los resultados de NAEP:

- Descripción del reporte de resultados de NAEP en la era de la web (Zenisky, Hambleton y Sireci, 1999). En este documento, se describen las diferentes herramientas de las que dispone NAEP en su página web para la presentación y el análisis de resultados. Es un estudio descriptivo, no proporciona estadísticas de uso de la información. Sin embargo, apunta fortalezas de la NAEP, en particular que no se genera una única manera de presentar los resultados: los usuarios pueden decidir la información que desean analizar y la forma de presentación de los datos. Se confirma la necesidad de que haya estudios evaluativos sobre el uso de esta información por diferentes audiencias.
- Reportes basados en preguntas para tomadores de decisiones (VanWinkle, Vezzu y Zapata-Rivera, 2011). En este documento se presenta una revisión de la literatura sobre las necesidades de información de los tomadores de decisiones y un prototipo de reporte para tomadores de decisiones basado en preguntas. Los hallazgos de una evaluación de NAEP realizada en 1997 se utilizan como uno de los elementos para sustentar la necesidad de elaborar reportes basados en preguntas para tomadores de decisiones (Hambleton y Slater, 1997). En dicho reporte se afirma que para los tomadores de decisiones los reportes de resultados de las pruebas

son difíciles de entender, por su experiencia, su familiaridad con los reportes, su conocimiento de la estadística y su tiempo disponible. Por lo anterior, se proponía que se simplificaran estos reportes, se probaran en campo y el contenido se adecuara a distintos tipos de audiencias.

- Evaluación de NAEP 2009 (Buckendahl *et al.*, 2009a, 2009b). En la evaluación más reciente de NAEP se analizaron la claridad y la accesibilidad de la comunicación de los reportes y resultados a los usuarios. En esta evaluación se concluyó que las estrategias de comunicación de resultados de NAEP han evolucionado de “recolección de datos y análisis a utilidad [*usabilidad*]” (Buckendahl *et al.*, 2009a: 86). En las entrevistas y grupos de enfoque que condujeron, encontraron que hay percepción positiva sobre el sitio web, aunque la información no parece ser comprensible y útil para las diferentes audiencias a las que se dirige. Una de las recomendaciones principales que se derivaron de este estudio fue el desarrollo de un marco de validez en el que se definen claramente los usos pretendidos de los resultados de NAEP y las interpretaciones posibles de sus puntajes. También se recomendó que los procesos de producción de manuales y reportes técnicos se revisaran de tal manera que se pudiera asegurar que se ponen a disposición del público en un tiempo adecuado.
- Como uno de los insumos de esta evaluación se realizó un estudio del uso del sitio web de NAEP entre el 1 de marzo de 2005 y el 28 de febrero de 2006 (Buckendahl *et al.*, 2009b: 4, 101-105), mediante los registros hechos por la compañía Webtrends. En este estudio se encontró que una de las herramientas más utilizadas por los usuarios fueron los reportes estatales; también se encontró un uso creciente de las herramientas de exploración de preguntas particularmente en el caso de docentes. El comportamiento de las consultas se incrementa particularmente cuando se publican los reportes de las evaluaciones.
- El uso de los resultados de NAEP en los estados (Ha, 2006). Esta tesis doctoral investigó la percepción del personal educativo de los estados en cuanto a la utilidad de NAEP y la manera en la cual se utilizan los datos. La investigación se realizó mediante entrevistas, documentos y análisis de sitios web. Se encontró que el uso en los estados es limitado y se enfoca principalmente en la divulgación de la información.

#### 4.2.4. Opinión pública

La búsqueda de la opinión pública sobre los usos de resultados de NAEP se hizo considerando las siguientes palabras clave: *NAEP in the news; NAEP data use; questions tool use; state profiles in the news; NAEP data explorer; NAEP simple questions*.

La búsqueda arrojó resultados vinculados principalmente con la descripción de los hallazgos de la última evaluación de NAEP. Se encontró una sección del *Huffington Post* dedicada a NAEP (<http://www.huffingtonpost.com/news/naep/>) concentrada en la descripción y el análisis de los resultados, pero no en las herramientas y prácticas de uso de la información de la evaluación.

Otro grupo de resultados que se desprendió de la búsqueda estuvo relacionado con los usos inadecuados de los resultados de NAEP (*NAEP data misuse*). Se enfatiza que los cambios en dichos resultados no pueden atribuirse a políticas educativas específicas, y además la necesidad de valorar los resultados de forma integral y no sólo enfocarse en asignaturas específicas o grados para

hacer públicos logros estatales, en los niveles más bajos de desempeño, o en el ordenamiento de los resultados por estados, entre otros usos inadecuados.

#### **4.2.5. Ejemplos de materiales de apoyo**

En este caso, los ejemplos de diversos materiales de apoyo utilizados en los Estados Unidos se encuentran en el texto del apartado anterior.

# 5

## El caso de PISA for Schools

---

### 5.1. Características de la evaluación a gran escala

#### 5.1.1. Introducción

PISA-Based Test for Schools, OECD Test for Schools, PISA for Schools, o también llamada PISA para Centros Educativos (en España), completó un ciclo de aplicación en el 2014 (Rutkowski, Rutkowski y Plucker, 2014). Cuatro son los países que actualmente han participado en pruebas de pilotajes de la versión del Programa Internacional de Evaluación de los Alumnos (PISA, por sus siglas en inglés) para escuelas: los Estados Unidos, Canadá, el Reino Unido y España (OECD, 2015a).

Estas pruebas derivan de la prueba general de PISA, la cual inició sus aplicaciones a partir del año 2000 con diferentes países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) (Wagemaker, 2013). Un desarrollo intermedio de PISA for Schools lo constituyen las pruebas International Student Assessment (ISA), realizadas por el Australian Council for Educational Research (ACER) (Gil y Sanz, 2014b), las cuales emplean el marco de PISA en la evaluación de Lenguaje, Matemáticas y Ciencias, para evaluar escuelas internacionales con un currículo internacional o con un foco internacional (ACER, 2015). Esta última sería el antecedente más cercano a PISA for Schools, debido a que también emplea un diseño para producir resultados a nivel de escuelas.

El objetivo general de PISA es proporcionar información sobre la población de estudiantes de cada país que se encuentra cerca del final de la educación obligatoria (15 años de edad), respecto del nivel de conocimientos y competencias esenciales para la participación en sociedades modernas. PISA for Schools comparte el mismo contenido, pero cambia el nivel inferencial. Esta evaluación brinda información acerca de los resultados de escuelas en particular, en relación no sólo a estándares nacionales, sino también en relación a los sistemas educacionales de mayor desempeño y las escuelas de mayor desempeño de otras regiones (OECD, 2012b).

En resumen, a diferencia de la prueba PISA tradicional, la cual está diseñada para brindar información para caracterizar los resultados de la población de estudiantes de 15 años de un país, PISA for Schools está diseñada para producir resultados para los centros educativos (MECD, 2013).

Los actores que se espera que hagan uso de esta información son principalmente, los directores de los establecimientos junto con sus profesores. No obstante, también se concibe que los usuarios pueden estar en otro orden jerárquico, como un nivel de distrito (EdLeader21, 2014b) u otra forma de gobierno local.

El uso esperado de los resultados obtenidos comprende objetivos de investigación, *benchmarking* y orientación de los esfuerzos de mejora escolar (OECD, 2015b). El uso pretendido de los resultados es que los actores de la comunidad reflexionen y discutan los resultados obtenidos, para luego diseñar planes de mejora.

### 5.1.2. Marco administrativo/legal

La forma en que PISA for Schools opera depende del organismo que contrate el servicio. Esta contratación puede establecerse como un convenio/suscripción por parte de un estado federal (por ejemplo, en los Estados Unidos: Massachusetts y Connecticut), una región de un país (como Dubai, Shanghái), una red de escuelas, o una escuela en particular. Actualmente, este servicio se encuentra en operación en los Estados Unidos, Reino Unido y España. No obstante, ya hay negociaciones con otros países para hacer este servicio disponible en otras regiones (OECD, 2015b).

Este tipo de evaluación no es obligatorio para las escuelas (Gil y Sanz, 2014b). Por ejemplo, en el caso del Reino Unido, una escuela puede suscribirse a esta evaluación por 3 550 libras (5 400 dólares, aproximadamente). Para el caso de los Estados Unidos, Rutkowski (2014) estima el costo por estudiante en 150 dólares.

En la documentación disponible por parte del Ministerio de Educación, Cultura y Deporte de España (MECD) se señala que las aplicaciones de las pruebas serían realizadas por los proveedores de servicios de evaluación disponibles que pudieran asegurar junto con los establecimientos escolares los estándares de calidad y procedimientos esperados (MECD, 2013). Para el caso del Reino Unido, el proveedor nacional es la National Foundation for Educational Research (NFER), la cual sería la responsable de la aplicación y la administración de las pruebas (NFER, 2014).

En la documentación técnica de PISA for Schools (OECD, 2015d) a esta figura se le llama genéricamente el Proveedor Nacional de Servicios (National Service Provider), entendiendo que este proveedor varía por país o región.

Para ilustrar la complejidad de los procesos involucrados en el desarrollo, la aplicación y el procesamiento de los datos de PISA for Schools, éstos pueden extraerse desde los documentos de muestra para los establecimientos pilotos (OECD, 2012b). Por ejemplo, la administración y el cálculo de puntajes estuvieron a cargo de CIB/McGrawHill (bajo la guía del OECD Directorate for Education), con la asistencia adicional de la NFER. Mientras que el desarrollo de ítems e *item equating* estuvo a cargo del ACER.

En términos generales los usos no pretendidos por esta evaluación son los siguientes (Gil y Sanz, 2014b):

- No es una prueba obligatoria.
- No pretende encontrarse alineada al currículo nacional (aunque podría haber solapamiento).

- No pretende influir por sí misma en las prácticas de enseñanza.
- No es una alternativa a las evaluaciones nacionales.
- No es una herramienta para establecer *rankings* (por ejemplo, *school league tables*).

Según declara la OCDE, los datos producidos por PISA for Schools no son públicos, y no se harán públicos (OECD, 2012b). Queda en responsabilidad de los administradores de los centros educativos el cómo harán difusión de los resultados obtenidos por su establecimiento (Gil y Sanz, 2014b).

### 5.1.3. Procedimientos y metodología

La metodología empleada por PISA for Schools es de dos tipos. En esta sección se describe la metodología general empleada por PISA, y adicionalmente se describe la metodología empleada por PISA for Schools. Ambos procedimientos han de ser considerados para comprender cómo funciona esta evaluación. Esto es debido a que PISA for Schools evalúa establecimientos que luego se comparan con los resultados internacionales de las diferentes aplicaciones de PISA (OECD, 2015d). Por conveniencia, a la primera sección la llamaremos "PISA principal", y a la segunda la seguiremos llamando PISA for Schools.

Los instrumentos empleados en PISA for Schools son de tres tipos: las pruebas cognitivas, las cuales evalúan la aplicación de conocimiento de Matemáticas, Ciencias y Lenguaje; el cuestionario para estudiantes, que registra aspectos motivacionales de los estudiantes, así como aspectos del ambiente escolar, y finalmente el cuestionario de la escuela, que recaba información del ambiente escolar y características generales del establecimiento (OECD, 2015d). Los estudiantes seleccionados contestan las pruebas y el cuestionario de estudiantes, mientras que el cuestionario del establecimiento se espera que sea completado por el director del establecimiento. Estos tres instrumentos se auto aplican en formato de papel y lápiz (Gil y Sanz, 2014a).

Los cuestionarios y pruebas son desarrollados por parte de la OCDE y por parte de ACER (OECD, 2015d). Las pruebas aplicadas consisten en una serie de ítems agrupados en diferentes formas, los cuales se rotan aleatoriamente entre los estudiantes (por ejemplo, *booklet random design*). Este diseño permite minimizar la cantidad de ítems que un estudiante tiene que responder (a no más de dos horas de aplicación). En conjunto, este diseño cubre una buena cantidad de los tópicos a evaluar, a pesar de que los estudiantes no se enfrentan a todos los ítems de la prueba, y aun así se producen inferencias válidas al nivel pretendido (Rutkowski, González, Joncas y Von Davier, 2010).

El cuestionario para estudiantes contiene ítems preexistentes de las rondas de PISA 2003, 2006, 2009 y 2012. La selección de ítems de este instrumento, está motivada para permitir la comparabilidad con las rondas de PISA principal, así como por la relevancia que poseen algunos de estos factores con los resultados a nivel de escuela (OECD, 2015d).

Por su parte, el cuestionario del establecimiento tiene por objeto caracterizar al establecimiento en su conjunto. Éste existe en dos versiones: la forma abreviada y la forma extendida. La forma abreviada recoge información tal como: el número total de estudiantes del establecimiento, el tipo de escuela y su sector (por ejemplo, privada o pública), la ubicación de la escuela, entre otros. Sólo toma unos minutos completar esta forma por parte del establecimiento. La forma extendida, en cambio, tiene un tiempo de duración esperada de unos 30 minutos. Registra toda la información

anterior y además pregunta por: la estructura organizacional de la escuela; la demografía de los profesores y estudiantes; los recursos de la escuela; la instrucción, el currículo y las evaluaciones de la escuela; el clima escolar; y las políticas y prácticas de la escuela. Los Proveedores de Servicios Nacionales (National Service Providers), en conjunto con la OCDE, pueden elegir si aplicar la forma abreviada o la forma extendida de este cuestionario (OECD, 2015d).

El diseño muestral que emplea la prueba PISA principal es probabilístico, estratificado de dos etapas. Primero se selecciona un conjunto aleatorio de escuelas, y luego se selecciona un conjunto aleatorio de estudiantes de 15 años que se encuentra dentro de cada escuela, con base en un diseño estratificado. En términos generales, el diseño aspira a recoger al menos 35 estudiantes por establecimiento, y unos 150 establecimientos por país/región. Este diseño es empleado para obtener parámetros confiables a nivel de la población (OECD, 2014).

En contraste, debido a que el interés de PISA for Schools es producir inferencias a nivel de las escuelas, el muestreo de estudiantes en el interior de las escuelas es mayor en estas aplicaciones. Como ya se mencionó, PISA muestrea a unos 35 estudiantes por establecimiento en su aplicación tradicional (OECD, 2014); mientras que PISA for Schools requiere muestras de 75 estudiantes por establecimiento (Gil y Sanz, 2014b). Otras fuentes indican que el número de estudiantes recomendados asciende a 84 estudiantes por escuela (NFER, 2015), o a 85 (OECD, 2015c).

En PISA for Schools no se requiere un muestreo probabilístico de escuelas, sino que las escuelas adquieren o registran su participación en esta evaluación de forma voluntaria. No obstante, la selección de los 75 a 85 estudiantes de 15 años en el interior del establecimiento serían seleccionados al azar (OECD, 2015d).

Existen al menos dos fechas en las que una escuela puede participar: en primavera y en otoño, para el caso del Reino Unido. Si la prueba se deseara para primavera (marzo), ésta debería ser suscrita en noviembre del año previo (NFER, 2014). Se asume que se emplearían criterios similares para otras regiones o países diferentes al Reino Unido. Las aplicaciones nunca son realizadas al inicio exacto del año escolar, sino después de unos dos meses de iniciado (o no más de tres meses). Este formato de aplicación evita que los resultados sean menores a lo esperado debido a la falta de práctica de los estudiantes como efecto de las vacaciones (OECD, 2015d).

Para que una aplicación sea realizada, el establecimiento que suscribe el servicio designa un coordinador. Éste sería la contraparte directa del establecimiento con el Proveedor Nacional de Servicios. Por su parte, se espera que el Proveedor Nacional de Servicios contrate a los Administradores del test, los cuales tienen por responsabilidad ejecutar la aplicación en los establecimientos, velando por la uniformidad y la imparcialidad del proceso, de acuerdo con los estándares internacionales de PISA for Schools (OCDE, 2015d).

El Proveedor Nacional de Servicios monitorea la recepción de materiales por ser aplicados, y la correcta asignación de éstos a los estudiantes predefinidos en la muestra de cada establecimiento. Este actor es responsable de la recepción del material aplicado con las respuestas de los establecimientos, su codificación y corrección, y la construcción de las bases de datos. Dicho proveedor tiene como responsabilidad la contratación y la capacitación de profesionales competentes para llevar a cabo el proceso de codificación y corrección de preguntas abiertas. A su vez, se espera que

cuenta con estadísticas competentes para llevar a cabo las tareas de procesamiento de datos, construcción de las bases de datos y cálculo de los resultados de los establecimientos, siguiendo los estándares de la OCDE (OECD, 2015d).

La población objetivo de PISA for Schools son los estudiantes de 15 años del establecimiento (OECD, 2015d). Esta prueba centra su foco en los estudiantes de 15 años, no en el grupo del curso o nivel por un asunto de comparabilidad. No todos los países agrupan a sus estudiantes por nivel o grupo de forma consistente. Existen países donde los estudiantes tienen diferentes compañeros según la asignatura en la que se encuentran. Un ejemplo de este tipo es Singapur, el cual agrupa a sus estudiantes no sólo por edad, sino también por avance en el currículo. Una forma de hacer comparables establecimientos que agrupan a sus estudiantes de diversa forma es emplear un mecanismo de muestreo común, tanto para establecimientos que agrupan a sus estudiantes en un solo curso o nivel, como para los establecimientos que sólo los distinguen por grupos de edad. De esta forma, la selección de 85 estudiantes al azar de 15 años proporciona un marco común para realizar inferencias en la población de los estudiantes de 15 años en general.

PISA for Schools evalúa tres contenidos aplicados y competencias de Matemáticas, Ciencias, y Lenguaje (Gil y Sanz, 2014a). Los estudiantes contestan entre 40 y 50 preguntas en cada área aproximadamente (Matemáticas=40 ítems, Ciencias=54 ítems, Lenguaje=47 ítems), las cuales son preguntas aplicadas a escenarios nuevos. En cerca de la mitad de éstas el estudiante tiene que desarrollar su propia respuesta, mientras que la otra mitad la constituyen preguntas con respuestas de alternativas múltiples (OECD, 2012e).

## 5.2. Características de las prácticas de uso de los resultados de las evaluaciones

### 5.2.1. Introducción

El uso esperado de los resultados obtenidos comprende objetivos de investigación, *benchmarking* y para orientar los esfuerzos de mejora escolar (OECD, 2015b). En términos generales, el uso pretendido de los resultados se centra en estimular la reflexión y la discusión entre educadores locales para la mejora escolar (OECD, 2015d).

Entre los usos no pretendidos se encuentra el establecimiento de *rankings* y el remplazo a las evaluaciones nacionales. Explícitamente, se recomienda no emplear estos resultados para esos usos (Gil y Sanz, 2014b).

Los usuarios principales de los resultados producidos son los directores y profesores de escuelas. El prototipo de uso se ilustra en la escena en que la escuela que encuentra bajos resultados en lectura los explica por la lectura restringida de sus estudiantes. Después, consulta con escuelas de similares características: cómo ellas promueven la lectura. Y luego de esta consulta, cambia sus planes de mejora aumentando las oportunidades de lectura recreativa de sus estudiantes. Este escenario podría constituir el uso ideal de resultados (ver EdLeader21, 2014b).

Los usos esperados de este tipo de resultados no se encuentran insertados en un proceso de acreditación ni de rendición de cuentas. Se espera que los usos de estos resultados sean un insumo

para la toma de decisiones de los establecimientos, para construir sus planes de mejora, y mediante la comparación con establecimientos de similares características se adopten metas adecuadas y logrables. Y, por otro lado, se busca que posean información adicional, de tipo no escolástico, para también abordar otros aspectos de la experiencia escolar, como la motivación de sus estudiantes, la relación con sus profesores, y la relación entre directores y profesores.

### 5.2.2. Herramientas para apoyar el uso de las evaluaciones

El principal producto de PISA for Schools es el reporte de resultados. Éste es un documento de aproximadamente 160 páginas. Actualmente existen tres reportes de muestra del informe de resultados que se entrega a las escuelas. (OECD 2012b; 2012d, y 2012f).

Por su parte Fairfax County Public Schools, de los Estados Unidos (ver EdLeader21, 2014a), ha compartido en línea los reportes de resultados recibidos por algunos de sus establecimientos (OECD 2012a; 2012c, y 2012d).

A diferencia de los tres primeros reportes, los cuales son de muestra, los informes del condado de Fairfax incluyen resultados reales (ver sección de ejemplos de este documento, para direcciones de todos los reportes por escuela).

El contenido de estos informes es específico para la escuela evaluada. En ellos se comparan los resultados de dicha escuela respecto a otras privadas y públicas de la misma región, en contraste con escuelas y regiones de la aplicación piloto de PISA for Schools, así como con resultados de otras regiones presentes en las aplicaciones previas de PISA, considerando el nivel socioeconómico medio de los estudiantes (Gil y Sanz, 2014b; NFER, 2015). Por otro lado, también se incluyen resultados respecto del ambiente de aprendizaje y sobre la motivación de los estudiantes, según los cuestionarios complementarios a las pruebas (Gil y Sanz, 2014b). Finalmente, los informes también incluyen recomendaciones mediante ejemplos de políticas relevantes y prácticas empleadas en diferentes regiones del mundo, las cuales buscan estimular la reflexión y la discusión entre educadores locales para la mejora escolar (OECD, 2015d).

El reporte de resultados producidos proviene de un reporte genérico producido desde una plantilla. A esta plantilla, mediante un *software*, se le agregan los datos y resultados de la escuela evaluada. El *software*, por su parte, remplacea: títulos, medias reportadas, figuras, secciones del texto en que se menciona la escuela, la descripción de los resultados de las figuras y los anexos. Una vez producido el reporte mediante este proceso, es revisado en búsqueda de resultados extremos: intervalos de confianza más grandes de los esperados, y valores puntuales muy extremos que pudieran encontrarse en los límites de las figuras empleadas. Por otro lado, si el nombre de la escuela es muy largo para las secciones del texto, éste requiere ser abreviado de modo que quepa dentro de los límites esperados (OECD, 2015d: 75).

Los informes de resultados generados por PISA for Schools son responsabilidad de la OCDE (OECD, 2015d) y pueden ser adaptados según la región y el establecimiento evaluado. Dichas adaptaciones consideran las comparaciones de escuelas públicas frente a privadas según sea aplicable. Otras adaptaciones posibles refieren a las diferentes terminologías empleadas a lo largo del informe. Tales adaptaciones al reporte original son realizadas por la OCDE en conjunto con el Proveedor Nacional (OECD, 2015d). Por ejemplo, en el caso del Reino Unido, el proveedor es la

NFER (2015), mientras que en el caso de España el Proveedor del Servicio Nacional es el Instituto Nacional de Evaluación Educativa (INEE), del MECD de España (Gil y Sanz, 2014b).

Respecto de las traducciones de los informes originales, la OCDE indica que sólo se considerarán válidos los reportes originales en inglés (OECD, 2015d). En otras palabras, la OCDE sólo garantiza la calidad de los informes producidos previa traducción, y no así la de traducciones posteriores. No obstante, todas las que se hacen de los informes de resultados de PISA for Schools son revisadas por parte de la OCDE en conjunto con el proveedor nacional, para buscar discrepancias entre la versión en inglés y su traducción. Aun así, la OCDE indica que no verifica la calidad y la validez del texto traducido, y que esto queda bajo la responsabilidad del proveedor nacional.

Los usuarios esperados de los informes de resultados son primariamente los administradores de los centros educativos. Queda en manos de los directivos de los centros educativos difundir estos resultados con sus profesores, otros educadores, los padres y sus estudiantes (Gil y Sanz, 2014b).

Como ya se ha mencionado, el objetivo principal que busca satisfacer esta herramienta, es la promoción de la reflexión y discusión de los resultados presentes de la escuela en su conjunto (Gil y Sanz, 2014b). La prueba aplicada evalúa si los estudiantes del establecimiento en promedio pueden no sólo reproducir conocimiento, sino adicionalmente aplicar este conocimiento a situaciones no familiares, tipos tanto escolar como no escolar (Gil y Sanz, 2014a).

El informe producido es una herramienta que le permite a los establecimientos compararse con escuelas de su propia región, así como con escuelas de otros países. En otras palabras, permite que las escuelas se comparen con sus pares, así como con otras escuelas en sistemas educativos líderes en el mundo (NFER, 2014).

PISA for Schools provee información a la escuela sobre el nivel estimado de cómo los estudiantes aplican conocimientos de Lenguaje, Matemáticas y Ciencias a problemas no familiares. Luego, estos niveles son comparados con escuelas de similares características dentro de la región de la escuela evaluada y con respecto al resto de las escuelas. Complementariamente, estos resultados también son comparados con otras regiones: como los resultados en el *top del ranking*, como Shanghái, así como con otras regiones. Finalmente, la herramienta de difusión y resultados posee recomendaciones y ejemplos de prácticas recogidos a lo largo del tiempo, desde diferentes escuelas y regiones por parte de PISA.

Como también ya se ha señalado, se espera que los directores, profesores, estudiantes y padres en conjunto participen en reuniones de reflexión sobre los resultados obtenidos (Gil y Sanz, 2014b).

Frente a este desafío, hay ejemplos de apoyos externos realizados por otras entidades diferentes a la OCDE y a las escuelas, que han participado en este proceso. Uno de estos ejemplos es el OECD Test for Schools Implementation Toolkit (EdLeader21, 2014b). Éste consiste en un documento complementario al reporte para escuelas en el que se espera resolver preguntas tales como:

- ¿Cómo pueden el distrito y sus directores promover un uso efectivo de PISA for Schools?
- ¿Cuáles son las tácticas más efectivas para administrar test el en la escuela, y comunicar a la comunidad acerca del mismo?
- ¿Cómo se puede sacar el mejor provecho al informe de resultados?

- ¿Cuáles son las estrategias más efectivas contenidas en el informe, para promover la mejora de aprendizaje de los estudiantes?
- ¿Qué fuentes accesibles hay para profundizar el conocimiento acerca de este proceso (listas, formas, preguntas frecuentes)?

La accesibilidad de los reportes de resultados es mayor a la que poseen los resultados internacionales de PISA reportados por la OCDE. Son menos técnicos, y aun así bastante completos en términos de metodología y descripción de procedimientos. No obstante, la accesibilidad de la información puede ser un desafío para los centros escolares. Prueba de esto son las iniciativas complementarias que acompañan a esta evaluación, como la ya mencionada Edleader21 (<http://edleader21.com/>), que proporciona una herramienta de trabajo específica para la implantación de PISA for Schools, tanto en sus etapas previas como en el trabajo posterior a la recepción de los resultados.

### 5.2.3. Uso de los resultados

Los diferentes usos de los resultados pueden ser clasificados por principio, entre los usos pretendidos y los usos no pretendidos (Rossi, Freeman y Lipsey, 1999). PISA for Schools aspira a que los resultados de cada establecimiento evaluado promuevan la discusión y la reflexión de las prácticas escolares presentes (OECD, 2015c).

Explícitamente, PISA for Schools no pretende establecer un *ranking* entre establecimientos (Gil y Sanz, 2014a); sin embargo, no puede evitar que los usuarios de los resultados sí lo hagan, gracias a la información comparativa que reciben. Por ejemplo, durante la aplicación de la prueba piloto el distrito de Blue Valley School, de Kansas, Estados Unidos, obtuvo de los mejores resultados de Kansas, muy cerca de los obtenidos por Finlandia, posicionando al distrito en el séptimo lugar de Matemáticas y el cuarto en Lenguaje. Durante las aplicaciones de 2013 y 2014 de PISA for Schools el distrito arengó a sus estudiantes a través de redes sociales como con el *hashtag* “#beat-finland” (Russo, 2014). Comparaciones competitivas entre Massachusetts y Connecticut ya han sido documentadas para las rondas de PISA 2012, y esto parece ser una práctica común de las evaluaciones subnacionales (Engel y Frizzell, 2015). Los usuarios de los resultados no sólo los emplean para compararse (por ejemplo, ver diferencias y similitudes), sino también para saber si están por sobre los resultados de otros.

PISA for Schools, como proceso de evaluación de escuelas, lleva muy poco tiempo como para poseer evaluaciones sobre el uso de sus resultados (Engel y Frizzell, 2015). Sin embargo, Edleader21 (2014b), en su herramienta de apoyo para PISA for Schools recopila seis estudios de caso donde se ejemplifican usos más o menos ideales de los resultados. Éstos son ejemplos de las escuelas y distritos que participaron en la aplicación piloto de PISA for Schools:

- El primer estudio de caso es sobre liderazgo de los profesores. Ilustra cómo PISA for Schools es adoptado fácilmente si la iniciativa de participar en la evaluación surge de los profesores y no del distrito. En términos generales, se espera que los profesores se opongan a involucrar a la escuela en otra evaluación adicional. Mientras que cuando la iniciativa comienza desde los profesores, no hay oponentes.

- El caso dos ilustra cómo la escuela de Chantilly High School revisa su plan de mejora, producto de los resultados recibidos. A esta escuela, si bien en las evaluaciones generales del estado no aparecía particularmente mal en lectura, la realimentación recibida por medio del reporte de resultados de PISA for Schools le permitió poner el foco en un aspecto particular: 44% de sus estudiantes eran lectores profundos, pero de contenido restringido (OECD, 2012a). Esto quiere decir que los estudiantes, si bien leen y comprenden, no leen por gusto, y sólo se concentran en los contenidos a leer que son obligatorios de las tareas escolares. Debido a estos resultados, el equipo directo de Chantilly School cambió sus prácticas y planes de mejora para promover la lectura motivada intrínsecamente mediante la inclusión de lectura recreativa (ver min. 20 de EdLeader21, 2014a).
- El caso tres, de Mount Vernon High School, es similar al caso dos, donde frente a un diagnóstico similar sobre el perfil lector de sus estudiantes (OECD, 2012c), la escuela también adopta un plan de mejora con el fin de aumentar las oportunidades de lectura de los estudiantes.
- El caso cuatro ilustra cómo el director de la escuela Oakton High School, la cual obtiene de los mejores resultados en el estado de Virginia (OECD, 2012d), incluso emplea la evaluación de PISA for Schools para metas relativas a su cargo, como mejorar su relación con los profesores.
- El caso cinco ilustra cómo North Star Academy compara sus resultados con las de escuelas de similares características, para recoger retroalimentación sobre cómo mejorar. Los resultados de PISA for Schools mostraban cómo, si bien no tenían puntajes bajos en ciencias, los aspectos motivacionales de esta materia se encontraban por debajo a lo esperado, y esto lo relacionaron a la baja salida que tienen sus estudiantes a carreras de ciencias (como ciencias y tecnología). La tónica consiste en identificar las brechas y consultar con escuelas similares qué prácticas emplean, para obtener mejores resultados.
- Finalmente, el caso seis ilustra brevemente la ventaja de una evaluación que proporciona comparaciones con otras escuelas de la región, así como de otros lugares del mundo. Esto es particularmente relevante para escuelas con una importante composición de estudiantes nacidos en el extranjero, ya que los resultados de la escuela son informativos no sólo respecto a los estándares nacionales, sino también con respecto a los resultados de otros países.

#### 5.2.4. Opinión pública

Para evaluar la presencia en los medios de PISA for Schools se realizaron búsquedas de las siguientes palabras claves: PISA-Based Test for Schools, OECD Test for Schools, PISA for Schools y PISA para Centros Educativos. Primero, en el sitio Socialmention,<sup>1</sup> que busca en twitter, blogs, noticias, y otras fuentes. No obstante, debido a lo nuevo de esta iniciativa (Engel y Frizzell, 2015), la presencia detectada por este medio es muy baja (sólo diez menciones, ocho neutras y dos positivas, para OECD Test for Schools, y ocho menciones para PISA for Schools). Una de las fuentes de difusión de PISA en los medios más visitadas es la charla en *TED talk*,<sup>2</sup> de Andreas Schleicher, coordinador

<sup>1</sup> <http://socialmention.com/>

<sup>2</sup> <https://www.youtube.com/watch?v=7Xmr87nsl74>

de PISA, con 40 855 visitas a la fecha. Ésta cuenta con 700 evaluaciones positivas (97%) y sólo 20 evaluaciones negativas (3%).

El contenido de las críticas que reciben estos contenidos en los medios es diverso, pero puede ser resumido. PISA for Schools está íntimamente asociada con PISA principal. Esto tiene como consecuencia, que quienes ya confían en PISA también confían en esta iniciativa. Sin embargo, para los críticos de PISA, esta otra iniciativa será tan cuestionable como aquélla, o más. De este modo, para comprender las posibles críticas que PISA for Schools enfrenta, vale la pena revisar las críticas que recibe PISA principal en general, y describir cómo PISA for Schools enfrenta estas críticas.

Las críticas principales que recaen sobre PISA en general pueden ser resumidas (Rutkowski y Rutkowski, 2014) en: la promoción de la agenda del mercado libre, una visión estrecha del sistema educativo, y un sesgo hacia los intereses de los países de la OCDE. Existe una oposición aún más amplia, que se opone a cualquier forma de evaluación estandarizada (como *large scale assessment*) y a cualquier forma de *ranking* de escuelas. PISA for Schools no puede hacerse cargo de todas estas críticas. Sin embargo, se cuida parcialmente de ellas sosteniendo explícitamente que no es una evaluación para establecer *rankings* (Gil y Sanz, 2014b).

Una de las críticas fundamentales que recibe PISA recae sobre el supuesto de comparabilidad entre los países y regiones que evalúa. El problema de comparabilidad posee dos aspectos: uno técnico y otro más práctico. El aspecto más práctico consiste en cuestionar el supuesto de la reingeniería de los resultados de la efectividad escolar (Lauder, Jamieson y Wikeley, 1998). La idea consiste en que una vez que se ha identificado quiénes obtienen el más alto desempeño académico, cualquier otro establecimiento podría adoptar las prácticas que éstos emplean para producir los mismos resultados. Gran parte de los ejemplos de usos de resultados de EdLeader21 (2014a) ilustran este principio entre escuelas. No obstante, suena inadecuado comparar países de gran diversidad sociodemográfica y contextual como los Estados Unidos, Rusia y México con países como Lichtenstein (Weiss, 2015), para adoptar prácticas ajenas aisladas de su contexto. La diversidad contextual entre países es un moderador importante de los resultados académicos de los estudiantes (Caro y Lenkeit, 2015). Interpretar los resultados logrados ignorando las condiciones socioeconómicas de los estudiantes, establecimientos y regiones puede llevar a establecer metas de logro inadecuadas para los establecimientos. De este modo, las evaluaciones nacionales que varios países ya poseen, como en el caso de la NAEP en los Estados Unidos, ofrecen una mayor ventaja para brindar información de resultados entre establecimientos de características similares (Rutkowski, Rutkowski y Plucker, 2014).

Frente a esta debilidad, la ventaja que posee PISA for Schools consiste en mostrar resultados comparables al establecimiento evaluado dentro de su propia región, como respecto a otras regiones y países con características similares. No obstante, la efectividad de esta característica depende de que los usuarios de los resultados puedan interpretar los resultados brutos como los resultados corregidos por nivel socioeconómico u otros factores relevantes, y así establecer comparaciones adecuadas e inferencias al respecto.

El aspecto más técnico por el cual los resultados de PISA son cuestionados recae en las fuentes de error que afectan a los resultados producidos: el error de muestreo y el error de medición. El primero consiste en juzgar si efectivamente los resultados publicados son representativos de la población de la que hacen inferencias. Uno de los casos más emblemáticos lo constituyen los

resultados de Shanghái en PISA 2012 (Loveless, 2001; Rutkowski y Rutkowski, 2014). Los resultados de Shanghái en PISA 2012 corresponden al primer lugar entre todas las regiones evaluadas. Sin embargo, la tasa de respuesta es de 79% de la población total de estudiantes de 15 años de la muestra pretendida (Rutkowski y Rutkowski, 2014). Por su parte, Loveless (2001) argumenta que la muestra de Shanghái es sistemáticamente sesgada, debido a que los estudiantes que asisten a las escuelas de Shanghái se encuentran preseleccionados por la política del *Hakou act*. Esta política, segrega a los estudiantes de origen inmigrante, quienes a pesar de atender la primaria en Shanghái, por ley son forzados a volver a las villas rurales de origen de sus padres. Como consecuencia, los estudiantes evaluados en Shanghái poseen una representatividad cuestionable de la población total de los estudiantes de 15 años, y además, el proceso de selección de estudiantes favorece a los estudiantes con mejores condiciones socio económicas y potencialmente mejores resultados (Loveless, 2001). Rutkowski, Rutkowski y Plucker (2014) comparan los resultados de los estudiantes con mejor desempeño de Shanghái y Massachusetts en PISA 2012, y muestran que las diferencias entre regiones, considerando estos ajustes, no son de gran tamaño.

Existen otros ejemplos documentados de sesgo muestral, como la participación de escuelas en Reino Unido (Schnepf, Durrant y Micklewright, 2014). En este estudio los autores evalúan si las características de las escuelas que participan y las escuelas que son elegidas en reemplazo de las que se rehúsan a participar poseen características que puedan sesgar los resultados. Los autores encuentran que las escuelas de remplazo se diferencian de las anteriores en sus características sociodemográficas, favoreciendo resultados más positivos.

El error de muestreo es difícil de manejar por parte de los usuarios, en ausencia de la manipulación de datos. El problema reside en que las muestras empleadas de cada país y región pueden ser más o menos representativas. Algunas correcciones pueden ser introducidas a los cálculos para obtener resultados menos sesgados, o que al menos se corrijan por estas diferencias. No obstante, PISA for Schools, no provee de datos a los centros evaluados, y la interpretación de resultados se encuentra restringida a los resultados provistos en los informes. De esta forma, PISA for Schools no puede corregir resultados sesgados producto del sesgo de muestreo sobre regiones y países de rondas pasadas.

El segundo aspecto de cautela, con los resultados de PISA for Schools, consiste en los problemas asociados al error de medición. Esto consiste en juzgar si las diferencias observadas entre establecimientos pueden ser atribuidas con exclusividad a lo que los establecimientos hacen y no a otras fuentes, tanto del instrumento mismo, como de características de los estudiantes. Esta segunda forma de error, referida al sesgo de los estimados, es más fácil de ejemplificar. Caro y Lenkeit (2015) muestran cómo el orden de los *rankings* varía entre países, cuando se corrigen los resultados por nivel socioeconómico.

PISA for Schools, por su parte, proporciona en su reporte de resultados bandas de comparación entre establecimientos que poseen características similares, para promover la comparación con otros centros que quizá sean referentes más válidos. Sin embargo, el problema de medición es más profundo, y refiere a la interacción entre el instrumento empleado y los sujetos evaluados. Mínimas variaciones en los conjuntos de ítems empleados pueden producir resultados diferentes atribuibles a los ítems empleados, y no a la habilidad de los sujetos evaluados (Rutkowski y Rutkowski, 2014). Este tipo de error no es fácil de aislar, y no es posible controlarlo sin manipular los datos recogidos.

En otras palabras, los usuarios de los resultados son potencialmente ciegos a esta debilidad, y PISA for Schools sólo puede tratar de corregir este aspecto antes de la construcción de informes. En términos sustantivos, esta fuente de error debería ser menor que las diferencias por sesgo muestral.

### 5.2.5. Ejemplos de materiales de apoyo

Los siguientes ejemplos son extraídos de los resultados reales de la escuela Chantilly High School, del condado de Fairfax, de los Estados Unidos (OECD, 2012a). La figura 7 muestra la comparación de los resultados de la escuela Chantilly High School a nivel local e internacional. La figura 8 enlista los perfiles de lectores considerados en PISA for Schools y una breve definición de los mismos. Finalmente, en la figura 9 se presenta el porcentaje de porcentaje de lectores en la escuela y a nivel nacional en cada uno de los perfiles de lectores.

Otras direcciones a materiales de apoyo disponibles en la web son las siguientes:

- EdLeader21 Toolkit de apoyo para la implantación de PISA for Schools: <http://edleader21.com/index.php?pg=27>
- Video conferencia con EdLeader21, que explicando el *toolkit* de apoyo para PISA for Schools: <http://edleader21.com/index.php?pg=33&id=29>
- Páginas de apoyo de Fairfax County Public Schools, Estados Unidos: [http://www.fcps.edu/pla/ost/pisa/pisa\\_2013\\_14\\_index.shtml](http://www.fcps.edu/pla/ost/pisa/pisa_2013_14_index.shtml); [http://www.fcps.edu/pla/ost/pisa/pisa\\_2012\\_13\\_index.shtml](http://www.fcps.edu/pla/ost/pisa/pisa_2012_13_index.shtml)

**Figura 7** Comparación de los resultados de la escuela, a nivel local e internacional



Fuente: OECD, 2012a. Traducción libre.

**Figura 8** Perfiles de lectores en PISA for Schools

	Superficial	Profundo
Ampio	<p><b>Lectores superficiales y amplios</b></p> <p>Estos estudiantes tienen niveles bajos de conocimiento sobre estrategias para entender, resumir y recordar información, pero leen una amplia variedad de materiales regularmente, incluyendo libros de ficción y no ficción. En los Estados Unidos, 7% de los estudiantes de 15 años son lectores superficiales y amplios.</p>	<p><b>Lectores profundos y amplios</b></p> <p>Estos estudiantes tienen altos niveles de conocimiento sobre estrategias efectivas de aprendizaje y además leen todo tipo de materiales por placer, incluyendo libros de ficción y de no ficción. En los Estados Unidos 19% de los estudiantes son lectores profundos y amplios.</p>
Estrecho	<p><b>Lectores estrechos y estrechos</b></p> <p>Estos estudiantes tienen niveles bajos de conocimiento sobre estrategias efectivas de aprendizaje y sus hábitos de lectura son estrechos en el sentido de que no leen una amplia variedad de materiales, aunque leer regularmente por placer. Este perfil agrupa a 6% de los estudiantes en los Estados Unidos.</p>	<p><b>Lectores profundos y estrechos</b></p> <p>Los estudiantes en este perfil también tienen altos niveles de conocimiento sobre estrategias efectivas de aprendizaje, pero sus hábitos de lectura son más estrechos que los de los lectores profundos y amplios. Este perfil agrupa a 11% de los estudiantes.</p>
Altamente restringido	<p><b>Lectores superficiales y altamente restringidos</b></p> <p>Los estudiantes en este perfil tienen bajos niveles de conocimiento sobre estrategias efectivas de aprendizaje y pasan poco tiempo leyendo por placer cualquier clase de materiales impresos. En los Estados Unidos 20% de los estudiantes son superficiales y altamente restringidos.</p>	<p><b>Lectores profundos y altamente restringidos</b></p> <p>Estos estudiantes tienen conocimiento de estrategias efectivas de aprendizaje, pero no leen con regularidad ningún tipo de materiales impresos por placer. Con 37%, este perfil agrupa al mayor porcentaje de estudiantes en los Estados Unidos.</p>

Fuente: OECD, 2012a. Traducción libre.



## Síntesis y conclusiones

---

En este informe se presenta una descripción sistemática de cinco conjuntos de prácticas internacionales en materia de uso de los resultados de evaluaciones a gran escala para dar realimentación a escuelas. El objetivo general de esta descripción fue identificar prácticas de posible implantación en el contexto de las pruebas del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). Los casos analizados son Alemania, Chile, España, Estados Unidos y el programa PISA for Schools, de la OCDE. La selección de estos casos se hizo obedeciendo a una serie de criterios descritos en el apartado dedicado a la metodología, así como con base en la disponibilidad de información para llevar a cabo el análisis propuesto.

Una primera conclusión de este estudio es precisamente la necesidad de llevar a cabo un proyecto adicional con la finalidad de desarrollar un sistema para el uso de información derivada de las pruebas PLANEA por parte de las escuelas. Dicho proyecto retomaría las acciones que en este sentido considera el documento rector de PLANEA (INEE, 2015a), además de las experiencias y ejemplos aquí recabados para desarrollar un sistema de uso de información especialmente diseñado para el contexto mexicano. Para alcanzar este objetivo, sería muy importante que el desarrollo o diseño de dicho sistema se realizara a partir del diálogo con los diferentes órganos colegiados del Instituto Nacional para la Evaluación de la Educación (INEE), así como con los diferentes actores involucrados en estos procesos —por ejemplo, la Secretaría de Educación Pública (SEP), las autoridades educativas estatales, las comunidades escolares, los docentes, los alumnos, los padres de familia, etcétera—. A grandes rasgos, este proyecto podría incluir las siguientes etapas:

- Evaluación de necesidades y expectativas de los diferentes grupos involucrados.
- Diseño de un sistema que empate (en medida de lo posible) los objetivos y características de PLANEA con las necesidades y expectativas de los actores involucrados.
- Diseño de un plan para la implementación del sistema que contenga las siguientes fases:
  - Piloteo.
  - Ajustes y rediseño.
  - Implementación definitiva.
- Evaluación y mejoramiento continuos.

Dicho lo anterior, a continuación se enlista una serie de estas prácticas que, de acuerdo con la información recabada, pueden considerarse exitosas en los contextos particulares en los que se utilizan.

## Vinculación entre los objetivos de las evaluaciones y los usos de la información que generan

Como lo indica el documento rector de PLANEA (INEE, 2015a: 27):

Una de las preocupaciones fundamentales de PLANEA son las necesidades de los distintos tipos de usuarios de todos los niveles del sistema educativo que se espera aprovechen los resultados de la evaluación para enriquecer sus decisiones. Estas decisiones van desde el diseño de políticas, planeación, gestión, monitoreo y rendición de cuentas, hasta la mejora de la enseñanza en los centros escolares.

Con respecto a la mejora de la enseñanza en los centros escolares, PLANEA considera la elaboración de Reportes de la Evaluación del Logro referida a centros escolares (INEE, 2015a: 27). Estos reportes tienen diferentes niveles de desagregación: dirigidos a las autoridades locales, a los supervisores escolares y a la comunidad escolar.

Estos planes para la difusión y uso de la información generada por PLANEA por parte de la comunidad escolar coinciden en términos generales con las prácticas analizadas en este documento. Sin embargo, la revisión de estas prácticas reveló algunos detalles que podrían fortalecer los planes señalados. Por ejemplo, PLANEA establece también como uno de sus propósitos "Ofrecer información pertinente, oportuna y contextualizada a las escuelas y a los docentes, que ayude a mejorar sus prácticas de enseñanza y el aprendizaje de todos sus estudiantes" (INEE, 2015a: 11).

En este sentido las prácticas analizadas consideran, además de las medidas consideradas en el documento Rector de PLANEA (por ejemplo, presentación de resultados siempre de manera contextualizada, la indicación explícita de no utilizarlos para elaborar *rankings*, etcétera), al menos dos puntos importantes:

- La elaboración de informes provisionales y finales. Es claro que el procesamiento y el análisis de este tipo de bases de datos es complejo y requiere de muchos días, semanas e incluso meses de trabajo. Por ello, las prácticas actuales en Chile, España y Alemania toman en cuenta la elaboración y la difusión de informes en dos tiempos. Un primer informe de carácter "provisional" tiende a utilizar la información que es más fácil de procesar. Estos reportes son generales en su naturaleza, pero tienen la ventaja de ofrecer información útil y sobre todo oportuna para los centros escolares. Un segundo informe o informe final tiende a utilizar la totalidad de los datos recolectados y a proporcionar un panorama más detallado. El equipo que trabaja en PLANEA ha preparado y distribuido reportes contextualizados de la Evaluación del Logro referida a los Centros Escolares (ELCE) para cada una de las escuelas que participaron en este ejercicio con remarcable rapidez (la evaluación se aplicó en junio de 2015 y los reportes se distribuyeron en noviembre del mismo año) (INEE, 2015b; SEP, 2015). Se considera, no obstante, que una revisión detallada de los contenidos de los informes preliminares y finales de Chile, Alemania y España podría ofrecer elementos para optimizar el contenido de los reportes

que actualmente se ofrecen a las escuelas mexicanas, y para evaluar la posibilidad y la utilidad de desarrollar reportes más detallados en un segundo momento.

- De manera adicional al desarrollo de materiales de apoyo para la interpretación de los informes (como también lo establece PLANEA), España también considera el envío de personal capacitado técnicamente a los centros escolares para explicar/demostrar la interpretación de informes escolares y de alumnos. El INEE de España tiene como una práctica establecida enviar a técnicos a explicar en ponencias monográficas el modo de interpretar estos informes. Con ello se fomenta que se usen de manera adecuada para generar planes de mejora de la calidad del sistema educativo de modo global.

### Actividades enfocadas a realizar la utilidad de estas evaluaciones para el quehacer diario de las escuelas

Como ya se mencionó, uno de los objetivos principales de PLANEA es ofrecer información a los docentes que ayude a mejorar sus prácticas de enseñanza, y el aprendizaje de todos sus estudiantes. Éste es un objetivo difícil de alcanzar debido a las características de las evaluaciones de gran escala. Como comenta Martínez Rizo *et al.* (2015), las pruebas de gran escala difícilmente pueden dar información sobre el aprendizaje de cada sustentante con el detalle necesario que permita poner en marcha acciones pedagógicas apropiadas. Esto debido a que estas pruebas se conforman con un número restringido de ítems, lo que limita la cobertura de contenidos del currículo, la precisión de los puntajes que se producen, y la demanda cognitiva general de las pruebas. Es claro que el diseño de PLANEA tiene procedimientos establecidos para contrarrestar estas limitaciones en la medida de lo posible, como las tres modalidades de aplicación (evaluación del sistema educativo, evaluación de logro en las escuelas y una evaluación diagnóstica censal coordinada por los propios docentes). Sin embargo, las prácticas analizadas en este documento ofrecen algunos ejemplos de acciones adicionales que podrían contribuir a resaltar la utilidad de la información producida por PLANEA para producir mejoras al nivel de la escuela:

- En este sentido el Instituto Nacional de la Evaluación Educativa de España ha venido realizando esfuerzos importantes. Ejemplos de ello son las actividades encaminadas al uso de ítems liberados de distintas evaluaciones de gran escala para realizar ejercicios y evaluaciones diagnósticas en el aula. Por ejemplo, el Massive Open Online Course (MOOC) sobre el uso de ítems liberados del Programa para la Evaluación Internacional de Alumnos y el Estudio de las Tendencias en Matemáticas y Ciencias (PISA y TIMSS, por sus siglas inglés, respectivamente) con fines didácticos. Los MOOC son una manera eficiente de involucrar grandes números de participantes a un bajo costo, además de fomentar la construcción del conocimiento a través de la experiencia colectiva, lo que a su vez promueve la apropiación de los contenidos por parte de los participantes. El MOOC que se menciona anteriormente tuvo más de diez mil inscripciones, de las cuales cerca de la mitad correspondió a maestros de América Latina (sin ser éstos el público objetivo), lo que da cuenta de la necesidad de este tipo de información y del buen recibimiento de estas iniciativas por parte de la comunidad de docentes.

- Otro ejemplo a destacar en el mismo sentido es el de los Estados Unidos. En su documentación, el National Assessment of Educational Progress (NAEP) especifica que los educadores pueden utilizar los marcos de referencia para diseñar evaluaciones y revisar el currículo; también se sugiere que las preguntas de respuesta construida y las guías para calificación podrían servir como modelo para desarrollar nuevas prácticas de evaluación en el aula (NCES, 2012). Estas actividades se pueden llevar a cabo por medio de herramientas especialmente diseñadas para tales fines (como Questions Tool, Interactive Items, Test Yourself, Item Maps) y podrían servir como complemento o funcionar en paralelo a la modalidad de PLANEA denominada como Evaluación Diagnóstica Censal.
- De manera similar a PLANEA, PISA for Schools aspira a que los resultados de cada establecimiento evaluado promuevan la discusión y la reflexión de las prácticas escolares presentes (OECD, 2015c). Ejemplos concretos de cómo hacerlo, de los que podría beneficiarse PLANEA (por ejemplo, para promover la discusión y reflexión de los reportes para las comunidades escolares) son los seis estudios de caso publicados por Edleader21.
- Un ejemplo más es el que da la Agencia de Calidad Escolar del estado federal de Nordrhein-Westfalen (QUA-LIS NRW), en Alemania. Este organismo ofrece amplia información en su página web para que los profesores puedan usar constructivamente los resultados obtenidos en las pruebas. Una vez más, aun cuando PLANEA ya tiene consideradas acciones similares, el modelo que utiliza QUA-LIS NRW podría contener elementos interesantes. Se destaca especialmente la claridad y el diseño holístico del sitio web. Las herramientas que en él se ofrecen se agrupan en secciones en el menú de resultados y la realimentación de la página web de esta agencia; cada una de ellas funciona como un insumo específico para alguna de las etapas que componen el modelo (ver figura 1). Ejemplos de estos insumos son la información para la correcta interpretación de los niveles de desempeño, para identificar resultados inesperados y posibles explicaciones para los mismos, y para hacer comparaciones con otras evaluaciones, así como materiales para la preparación de realimentación para los estudiantes y sus familias, y para el mejoramiento de las prácticas de aula. Si bien el material está bien organizado y es claro, contiene mucha información y bastante compleja, lo que supone un usuario familiarizado con este tipo de resultados y que además dispone del tiempo necesario para procesar y analizar la información.

## Uso de redes sociales

Un elemento sin duda interesante es el uso de las redes sociales, ya sea para promover o para criticar las acciones encaminadas al uso de las evaluaciones a gran escala para dar realimentación a las escuelas. En el caso de Chile, por ejemplo, la mayor parte de las críticas se formaron, distribuyeron y debatieron a través de las redes sociales (ver, por ejemplo, <http://www.altoalsimce.org>).

En el caso de España, resalta el uso intensivo de redes sociales (como twitter, YouTube, etcétera) y listas de correos electrónicos para informar a los actores involucrados acerca de novedades y eventos, e incrementar el impacto de dichas acciones. También se usa de manera importante para promover el debate sobre estos temas entre el público interesado (ver, por ejemplo, <http://blog.educalab.es/inee/>).

## Apoyos para el fomento de análisis secundarios

Al igual que todos los casos analizados, PLANEA considera poner las bases de datos de esta evaluación a disposición de universidades, asociaciones y público en general, de manera que sea posible realizar análisis complementarios a los que la SEP y el INEE aporten (INEE, 2015a: 28). Esto es sin duda importante e indispensable para promover el uso y el debate sobre la información a generarse, pero las prácticas que se analizan en este documento sugieren que no es suficiente. Algunas acciones descritas en los apartados anteriores de este documento son:

- Organización de ponencias de especialistas internacionales (investigadores y practicantes), y la difusión masiva de los videos del evento, como en el caso de España. Este tipo de eventos tienen el objetivo final de promover el uso de las bases de datos entre la comunidad científica para continuar trabajos relacionados con las evaluaciones educativas a gran escala.
- En España y en Alemania es habitual organizar congresos posteriores a la presentación oficial de los informes. En estos congresos se invita a docentes, equipos directivos, investigadores y administradores públicos. En Alemania la organización de estos congresos está a cargo de una asociación independiente (Asociación de Desarrollo Escolar Basado en Evidencia Empírica, EMSE). La EMSE está compuesta por 80 representantes de las distintas agencias de calidad escolar de los 16 estados federales, así como por miembros de instituciones académicas y universidades. La estructura de estos congresos es normalmente: presentación de informes por el organismo que lleva a cabo la evaluación, presentación de análisis secundarios por investigadores invitados, mesas redondas compuestas por administradores de políticas educativas, mesas redondas compuestas por escuelas con buenas prácticas, y, como eje central, la participación de los asistentes en debates sobre los temas tratados.

### Carga laboral y tiempo asignado para hacer uso de los resultados de las evaluaciones por parte de los docentes

Es bien sabido que el objetivo último de toda evaluación es la mejora de la calidad educativa, y que las evaluaciones de gran escala pueden contribuir a alcanzar este objetivo al proporcionar a los maestros referentes útiles para poner en perspectiva el nivel de aprendizaje de sus propios alumnos. Sin embargo, ésta es una actividad que requiere tiempo, compromiso y la adecuada capacitación técnica. En los casos analizados en este documento un tema recurrente fue la importancia de considerar el tiempo y los conocimientos técnicos que los docentes y directores de escuela necesitan para hacer uso adecuado de la información proveniente de las evaluaciones a gran escala y de sus herramientas de apoyo. Si bien se entiende que la asignación de cargas laborales a docentes y directores de escuela está fuera del área de influencia del INEE, las prácticas revisadas sugieren la existencia de otras alternativas para atender este punto.

Por ejemplo, en Alemania se han manifestado reparos concretos respecto de la sobrecarga que implica para los profesores el uso de esta información, así como dudas respecto de sus beneficios, considerando que es una iniciativa que implica costos elevados. En este respecto, en el estado alemán de Hessen (de manera similar a lo que hace el INEE de España) se decidió crear la figura del asesor de desarrollo de prácticas de aula, quien va a las escuelas a colaborar en la instalación

de capacidades en torno al uso de resultados de pruebas nacionales alemanas (VERA). Una evaluación preliminar de pequeña escala de esta iniciativa (Vogel, Achmetli, Blum y Krawitz, 2014) reveló que los resultados han sido positivos para el grupo de control en términos de un aumento en la aceptación de las pruebas y la percepción acerca de su utilidad con propósitos diagnósticos y como insumo para las prácticas docentes, conservándose, no obstante, la crítica acerca de la sobrecarga laboral que supone.

Adicionalmente, como se apunta en el punto 4 de estas conclusiones, los apoyos para el fomento de análisis secundarios podrían diseñarse de manera que contribuyan directamente a la formación de capacidad técnica entre los docentes y directivos de las escuelas.

### **Evaluar el uso de la información que se produce por parte de los diversos destinatarios**

La revisión y el seguimiento que considera PLANEA en su documento rector (INEE, 2015a: 30) es sin duda fundamental para asegurar la calidad y pertinencia de la información que producirá. Un ejemplo interesante en este sentido puede ser el que ofrece el caso chileno, en particular por las fuertes críticas que el Sistema de Medición de la Calidad de la Educación (SIMCE) ha recibido en los últimos años y, sobre todo, por la forma en que la Agencia de Calidad de la Educación ha reaccionado.

Como se describe en los apartados anteriores, en fechas recientes, en Chile se convocó a un Equipo de Tarea para la Revisión del SIMCE. Este equipo se creó en el contexto de diversos cambios en las políticas públicas de evaluación educacional (como la clasificación de escuelas acorde con la Ley de Aseguramiento de la Calidad) y una serie de manifestaciones críticas al SIMCE (ver, por ejemplo, <http://www.altoalsimce.org>) por parte de movimientos sociales y de un grupo de investigadores y académicos que firmaron una carta abierta para cuestionar varias características de la medición nacional, como también sus políticas asociadas. Este equipo, además de equipos independientes de investigadores, evaluó los usos intencionados y no intencionados de la aplicación y la comunicación de resultados de las pruebas SIMCE, y generó recomendaciones puntuales para su mejora (ver MINEDUC, 2015; Taut, Cortes, Sebastian y Preiss, 2009; y Manzi, Bogolasky, Gutiérrez, Grau y Volante, 2014).

## Referencias

- ACER. Australian Council for Educational Research (2015). ISA International Schools' Assessment Program. Camberwell, Australia: Australian Council for Educational Research. Recuperado de: [https://www.acer.edu.au/files/ISA\\_About.pdf](https://www.acer.edu.au/files/ISA_About.pdf)
- Agencia de Calidad de la Educación (2015). *Informe técnico 2013*. Santiago de Chile: Autor.
- Bauman, A. (2011, 7 de junio). Ein Test namens VERA [Un test llamado VERA]. *Die Zeit*. Recuperado de: <http://www.zeit.de/gesellschaft/schule/2011-06/leserartikel-schule-vera>
- Bravo, J. (2011). SIMCE: Pasado, presente y futuro del Sistema Nacional de Evaluación. *Estudios Públicos*, 123, pp. 189-211.
- Buckendahl, C. W., Davis, S. L., Plake, B. S., Sireci, S. G., Hambleton, R. K., Zenisky, A. L. y Wells, C. S. (2009a). Evaluation of the National Assessment of Educational Progress. Citeseer. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.183.8938&rep=rep1&type=pdf>
- Buckendahl, C. W., Davis, S. L., Plake, B. S., Sireci, S. G., Hambleton, R. K., Zenisky, A. L. y Wells, C. S. (2009b). *Evaluation of the National Assessment of Educational Progress. Study Reports*. Washington, D. C.: Office of Planning, Evaluation and Policy Development, US Department of Education. Recuperado de: <http://eric.ed.gov/?id=ED526784>
- Caro, D. y Lenkeit, J. (2015). Which Countries Punch above their Weight in Education Rankings? *The Conversation*. Recuperado de <https://theconversation.com/which-countries-punch-above-their-weight-in-education-rankings-49698>
- Elmore, R. (2008). *School Reform from the Inside Out: Policy, Practice and Performance*. Cambridge M.A.: Harvard Education Press.
- EdLeader21 (2014a). OECD Test for School Webinar. Recuperado de <http://edleader21.com/index.php?pg=33&id=29>
- EdLeader21 (2014b). OECD Test for Schools. Implementation Toolkit. Recuperado de: [http://www.oecd.org/pisa/aboutpisa/EdLeader21\\_OECD\\_TFS\\_Toolkit.pdf](http://www.oecd.org/pisa/aboutpisa/EdLeader21_OECD_TFS_Toolkit.pdf)
- Engel, L. C. y Frizzell, M. O. (2015, julio). Competitive Comparison and PISA Bragging Rights: Sub-national Uses of the OECD's PISA in Canada and the USA. *Discourse: Studies in the Cultural Politics of Education*, 6306, pp. 1-18.
- Gil, G. y Sanz, I. (2014a). PISA para Centros Educativos: El futuro de la evaluación del rendimiento de los centros. *Ruta Maestra*, 07, pp. 43-48. Recuperado de: [http://www.santillana.com.co/rutamaestra/revistas\\_pdf/ruta\\_maestra\\_v\\_007.pdf](http://www.santillana.com.co/rutamaestra/revistas_pdf/ruta_maestra_v_007.pdf)
- Gil, G. y Sanz, I. (2014b). PISA para Centros Educativos: Un instrumento para el fomento del éxito escolar de los centros. *Participación Educativa*, 3(4), pp. 29-37.

- Ha, J.-M. (2006). *The Use of NAEP Data in a State Context* (tesis de doctorado en ciencias de la educación). Oregon State University. Recuperado de: <http://ir.library.oregonstate.edu/jspui/handle/1957/3876>
- Hambleton, R. K. y Slater, S. (1997). Are NAEP Executive Summary Reports Understandable to Policy Makers and Educators? (CSE Technical Report, 430).
- Hein, A. y Taut, S. (2010). El uso de información evaluativa externa con fines formativos: El caso de establecimientos educacionales chilenos participantes de SEPA. *Revista Iberoamericana de Evaluación Educativa*, 3(2), pp. 160-181.
- Hessische Lehrkräfteakademie (2015). Lerstanderhebungen. Teilnahm und Termine [Mediciones de estado, participación y fechas]. Recuperado de: [https://la.hessen.de/irj/LSA\\_Internet?cid=052e80f8413c9fdeb1b8c8093152d5c5](https://la.hessen.de/irj/LSA_Internet?cid=052e80f8413c9fdeb1b8c8093152d5c5)
- Himmel, E. (1997). Impacto de los sistemas de Evaluación del rendimiento Escolar: El Caso de Chile. En B. Álvarez y M. Ruiz-Casares (Eds.), *Evaluación y Reforma Educativa: Opciones de Política*. PREAL.
- IES. Institute of Education Sciences (2015). NAEP Long-Term Trend Assessments. Recuperado de: <http://nces.ed.gov/nationsreportcard/ltt/>
- IES (2016). NAEP Long-Term Trend Assessments. Recuperado de: <http://www.nationsreportcard.gov/>
- INEE. Instituto Nacional para la Evaluación de la Educación (2015a). *Plan Nacional para la Evaluación de los Aprendizajes (PLANEA)*. Recuperado de: <http://planea.sep.gob.mx/content/general/docs/2015/PlaneaDocumentoRector.pdf>
- INEE (2015b). Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). Resultados nacionales 2015. 6° de primaria y 3° de secundaria. Lenguaje y comunicación. Matemáticas. Recuperado de: <http://www.inee.edu.mx/images/stories/2015/planea/final/fasciculos-finales/resultadosPlanea-3011.pdf>
- IQR. Institut zur Qualitätsentwicklung im Bildungswesen (2015). VERA—Ein Überblick [VERA—Una sinopsis]. Recuperado de: <https://www.iqb.hu-berlin.de/vera>
- KMK. Kultusministerkonferenz (2012). Vereinbarung zur Weiterentwicklung von VERA [Acuerdo acerca del desarrollo y proyecciones de VERA]. Recuperado de: <http://www.kmk.org/dokumentation/veroeffentlichungen-beschluesse/bildung-schule/qualitaetssicherung-in-schulen.html#c7076>
- KMK (2014). The Education System in the Federal Republic of Germany 2012/2013. A description of the responsibilities, structures and developments in education policy for the exchange of information in Europe. Recuperado de: <http://www.kmk.org/dokumentation/das-bildungswesen-in-der-bundesrepublik-deutschland/dossier-englisch/publikation-zum-download.html>
- KMK (2015). Bildungsstandards, ein Überblick [Estándares de enseñanza, una sinopsis]. Recuperado de: <http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/ueberblick.html>
- Lauder, H., Jamieson, I. y Wikeley, F. (1998). Models of Effective Schools: Limits and Capabilities. En R. Slee, G. Weiner y S. Tomlinson (Eds.), *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*. Londres: Falmer Press.
- Ley 20.529 (2011). Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización. Recuperado de: <http://www.leychile.cl/N?i=1028635&f=2015-05-05&p=>
- Loveless, T. (2001). How Well are American Students Learning. *The 2001 Brown Center Report on American Education*, 1(2). Recuperado de <http://www.brookings.edu/~media/research/files/reports/2001/9/education/09education.pdf>

- Manzi, J., Bogolasky, F., Gutiérrez, G., Grau, V. y Volante, P. (2014). Análisis sobre valoraciones, comprensión y uso del SIMCE por parte de directores escolares de establecimientos subvencionados. Informe preliminar FONIDE (no publicado).
- Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: Hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, 11(2), pp. 1-18.
- Martínez Rizo, F. et al. (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación* (Cuaderno de Investigación núm. 40). México: INEE.
- McLauchlan de Arregui, P. (2012). Difusión y uso de los resultados de evaluaciones educativas a gran escala en América Latina. En E. Martín y F. Martínez Rizo (Coords.). *Avances y Desafíos en la evaluación educativa*. Madrid: OEI/Fundación Santillana.
- MECD. Ministerio de Educación, Cultura y Deporte (2013). PISA para Centros Educativos (PISA for Schools). Recuperado de: <http://www.mecd.gob.es/dctm/inee/pisa-for-school/informacionbasicapisaforschools.pdf?documentId=0901e72b8162dd19>
- Meckes, L. (2003). El SIMCE: Su desarrollo y desafíos actuales. *Pensamiento Educativo*, 33, pp. 160-178.
- Meyer, T. (2014, diciembre). Evaluationsergebnisse zur Nutzung der VERA 8 Resultate in den Fachkonferenzen der Fächer Englisch und Französisch [Resultados de la evaluación del uso de los resultados de VERA en los consejos técnicos de las asignaturas de inglés y francés]. Trabajo presentado en el 19º seminario de la Asociación de Desarrollo Escolar Basado en Evidencia Empírica (EMSE), en Hamburgo, Alemania. Recuperado de: <http://www.emse-netzwerk.de/pmwiki.php/Main/Material>
- MINEDUC. Ministerio de Educación-Comisión para el Desarrollo y Uso del SIMCE (2003). *Evaluación de Aprendizajes para una Educación de Calidad*. Santiago de Chile: autor.
- MINEDUC-Equipo de Tarea para la Revisión del SIMCE (2015). *Hacia un sistema completo y equilibrado de evaluación de los aprendizajes en Chile*. Santiago de Chile: autor.
- NAGB. National Assessment Governing Board (s.f.). National Assessment Governing Board. Who we are. Recuperado de: <https://www.nagb.org/content/nagb/assets/documents/who-we-are/PDFforOverviewSection.pdf>
- NCES. National Center for Education Statistics (2009). An Overview of Procedures for the NAEP Assessment. US Department of Education. Recuperado de: <https://nces.ed.gov/nationsreportcard/pdf/about/2009493.pdf>
- NCES (2010a). An Introduction to NAEP. Recuperado de: <http://nces.ed.gov/nationsreportcard/pdf/parents/2010468.pdf>
- NCES (2010b). Information about NAEP for you. Recuperado de: <http://nces.ed.gov/nationsreportcard/infofor.aspx>
- NCES (2010c). Visión general de NAEP. Recuperado de: <https://nces.ed.gov/nationsreportcard/pdf/parents/2010457rev.pdf>
- NCES (2011). NAEP Assessment Sample Design. Recuperado de: [https://nces.ed.gov/nationsreportcard/tdw/sample\\_design/](https://nces.ed.gov/nationsreportcard/tdw/sample_design/)
- NCES (2012). Frequently Asked Questions. Recuperado de: <http://nces.ed.gov/nationsreportcard/faq.aspx#ques3>
- NCES (2013). An Overview of NAEP. US Department of Education. Recuperado de: [http://nces.ed.gov/nationsreportcard/subject/\\_commonobjects/pdf/2013455.pdf](http://nces.ed.gov/nationsreportcard/subject/_commonobjects/pdf/2013455.pdf)
- NCES (2014a). Field Staff Structure since 2002. Recuperado de: [https://nces.ed.gov/nationsreportcard/tdw/data\\_collection/staff\\_structure\\_since\\_2002.aspx](https://nces.ed.gov/nationsreportcard/tdw/data_collection/staff_structure_since_2002.aspx)
- NCES (2014b). NAEP Data Collection. Recuperado de: [https://nces.ed.gov/nationsreportcard/tdw/data\\_collection/](https://nces.ed.gov/nationsreportcard/tdw/data_collection/)

- NCES (2015a). How NAEP is Administered. Recuperado de: <https://nces.ed.gov/nationsreportcard/about/natadministered.aspx>
- NCES. (2015b). The history of NAEP contractors. Recuperado de: [https://nces.ed.gov/nationsreportcard/contracts/history.aspx#current\\_contracts](https://nces.ed.gov/nationsreportcard/contracts/history.aspx#current_contracts)
- NFER. National Foundatio for Educational Research (2014). NFER launches PISA-Based Test for Schools in England, Wales, Northern Ireland and the Channel Islands. Recuperado de: <http://www.nfer.ac.uk/about-nfer/media-and-events/nfer-launches-pisa-based-test-for-schools-in-england-wales-northern-ireland-and-the-channel-islands.cfm>
- NFER (2015). PISA-Based Test for Schools. Recuperado de <http://www.nfer.ac.uk/schools/pisa-based-test-for-schools/>
- OECD. Organisation for Economic Co-operation and Development (2012a). How your school compares internationally. OECD Test for Schools (Based on PISA) Pilot Trial. Chantilly High School. Recuperado de: [http://www.fcps.edu/pla/ost/pisa/school\\_level\\_reports/chantilly\\_hs.pdf](http://www.fcps.edu/pla/ost/pisa/school_level_reports/chantilly_hs.pdf)
- OECD (2012b). How your School Compares Internationally. OECD Test for Schools (Based on PISA) Pilot Trial. Herndon High School. Recuperado de: <http://www.oecd.org/pisa/aboutpisa/OECD%20Test%20for%20Schools%20-%20Herndon%20High%20School.pdf>
- OECD (2012c). How your School Compares Internationally. OECD Test for Schools (Based on PISA) Pilot Trial. Mount Vernon High School. Recuperado de: [http://www.fcps.edu/pla/ost/pisa/school\\_level\\_reports/mount\\_vernon\\_hs.pdf](http://www.fcps.edu/pla/ost/pisa/school_level_reports/mount_vernon_hs.pdf)
- OECD (2012d). How your School Compares Internationally. OECD Test for Schools (Based on PISA) Pilot Trial. Oakton High School. Recuperado de: [http://www.fcps.edu/pla/ost/pisa/school\\_level\\_reports/oakton\\_hs.pdf](http://www.fcps.edu/pla/ost/pisa/school_level_reports/oakton_hs.pdf)
- OECD (2012e). How your School Compares Internationally. OECD Test for Schools (Based on PISA) Pilot Trial. North Star Academy College Preparatory High School. Recuperado de: <http://www.americaachievers.org/docs/OECD/OECD-Sample-Test-For-Schools-Report.pdf>
- OECD (2012f). How your School Compares Internationally: PISA-Based Test for Schools Pilot Trial. The European School Culham. Recuperado de: <http://www.oecd.org/pisa/aboutpisa/PISA-Based-Test-for-Schools-European-School-Culham.pdf>
- OECD (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment* (OECD Reviews of Evaluation and Assessment in Education). París: autor.
- OECD (2014). *PISA 2012 Technical Report*. París: autor. Recuperado de: <http://www.oecd.org/pisa/pisaproducts/pisa2012technicalreport.htm>
- OECD (2015a). Country Information. PISA-Based Test for Schools. Recuperado de: <http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools-country-specific-information-and-global-learning-network.htm>
- OECD (2015b). PISA-Based Test for Schools—FAQ. PISA-Based Test for Schools. Recuperado de: <http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools-faq.htm>
- OECD (2015c). PISA-Based Test for Schools. Brochure. Recuperado de: <http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools.htm>
- OECD (2015d). PISA-Based Test for Schools. Technical Report. París: autor. Recuperado de: [http://www.oecd.org/pisa/aboutpisa/PfS\\_TechReport\\_CRC\\_final.pdf](http://www.oecd.org/pisa/aboutpisa/PfS_TechReport_CRC_final.pdf)
- QUA-LIS NRW. Qualitäts- und UnterstützungsAgentur- Landesinstitute für Schule Nordrhein Westfalen (2015) Rückmeldung, Reflexion und Adressaten der Ergebnisrückmeldungen [Feedback, Reflexión y destinatarios de los reportes de resultados]. Recuperado de: <http://www.schulentwicklung.nrw.de/vera3/lehrerinformationen/ergebnisse/>
- Rossi, P. H., Freeman, H. E. y Lipsey, M. W. (1999). *Evaluation: A Systematic Approach* (6a. ed.). Londres: SAGE Publications.

- Russo, A. (2014). Quietly, a New Test Gains Advocates. Harvard Graduate School of Education. Recuperado de: [http://hepg.org/hel-home/issues/30\\_3/helarticle/quietly,-a-new-test-gains-advocates](http://hepg.org/hel-home/issues/30_3/helarticle/quietly,-a-new-test-gains-advocates)
- Rutkowski, D. (2014, noviembre). The OECD and the local: PISA-Based Test for Schools in the USA. *Discourse: Studies in the Cultural Politics of Education*, 6306, pp. 1-17.
- Rutkowski, D., Rutkowski, L. y Plucker, J. A. (2014). Should Individual U. S. Schools Participate in PISA? *Phi Delta Kappan*, 96(4), pp. 68-73.
- Rutkowski, L., Gonzalez, E., Joncas, M. y Von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, 39(2), pp. 142-151.
- Rutkowski, L. y Rutkowski, D. (2014). OECD's Leaning Tower of PISA: Approach with Caution, Use with Care. European Commission. Recuperado de: <https://ec.europa.eu/jrc/sites/default/files/2014-pisa-leaning-iu.pdf>
- Sandoval-Hernández, A. y Neuschmidt, O. (2014). Lecciones de más de cinco décadas de estudios internacionales a gran escala sobre el logro educativo. *Organización y Gestión Educativa*, 22(6), 21-25.
- Schnepf, S. V., Durrant, G. B. y Micklewright, J. (2014). Which Schools and Pupils Respond to Educational Achievement Surveys? A Focus on the English PISA Sample (IZA Discussion Papers núm. 8411). Bonn, Germany. Recuperado de: <http://www.econstor.eu/bitstream/10419/101880/1/dp8411.pdf>
- SEP. Secretaría de Educación Pública (2015). Resultados PLANEA Básica 2015. Evaluación del logro referida a centros escolares. Recuperado de: [http://planea.sep.gob.mx/content/general/docs/2015/difusion\\_resultados/2\\_Planea\\_Basica\\_2015\\_ELCE.pdf](http://planea.sep.gob.mx/content/general/docs/2015/difusion_resultados/2_Planea_Basica_2015_ELCE.pdf)
- Spiegelonline, Schulspiegel (2014). Leistungsdruck an Schulen: Lehrer schimpfen über Vergleichstests [Presión por rendimiento en colegios: El profesorado critica fuertemente las pruebas comparativas]. *Der Spiegel*. Recuperado de: <http://www.spiegel.de/schulspiegel/vergleichsarbeiten-vera-an-schulen-lehrer-wettern-gegen-tests-a-967642.html>
- Taut, S., Cortés, F., Sebastian, C. y Preiss, D. (2009). Evaluating School and Parent Reports of the National Student Achievement Testing System (SIMCE) in Chile: Access, Comprehension, and Use. *Evaluation and Program Planning*, 39, pp. 129-137.
- Universidad Alberto Hurtado (2008). *Estudio evaluación de la Jornada de Análisis de Resultados SIMCE de 2007 con niveles de logro*. Santiago de Chile: CIDE/Universidad Alberto Hurtado.
- VanWinkle, W., Vezzu, M. y Zapata-Rivera, D. (2011). Question-based reports for policymakers (Research Memorandum, 11-16).
- Vogel, Achmetli, Blum y Krawitz (2014, diciembre). VELM-8 — ein Projekt zur Nutzung der Lernstandserhebungen für Unterrichtsentwicklung und Förderung [VELM-8, un proyecto sobre el uso de las mediciones de estado para desarrollo y promoción de la enseñanza]. Trabajo presentado en el 19º seminario de la Asociación de Desarrollo Escolar Basado en Evidencia Empírica (EMSE), en Hamburgo, Alemania. Recuperado de: <http://www.emse-netzwerk.de/pmwiki.php/Main/Material>
- Wagemaker, H. (2013). International Large-Scale Assessments: From Research to Policy. En L. Rutkowski, M. von Davier, y D. Rutkowski (eds.). *Handbook of International Large-scale Assessment: Background, Technical Issues, and Methods of Data Analysis* (pp. 11-36). Boca Raton, Londres y Nueva York: Chapman and Hall/CRC.
- Weiss, E. (2015). Making Sense of Assessments: State level NAEP Comparisons Offer Better Policy Guide Than International Assessments. Huffington Post. Recuperado de: [http://www.huffingtonpost.com/elaine?weiss/making?sense?of?assessments\\_b\\_8473156.html](http://www.huffingtonpost.com/elaine?weiss/making?sense?of?assessments_b_8473156.html)

Zenisky, A. L., Hambleton, R. K., y Sireci, S. G. (1999). At or Above Proficient: The Reporting of NAEP Results in the Internet Age. Center for Educational Assessment Report No. 710.

Recuperado de: [http://www.umass.edu/remf/docs/NCME09\\_ZHS\\_v3.pdf](http://www.umass.edu/remf/docs/NCME09_ZHS_v3.pdf)

ZEPF. Zentrum für Empirische Pädagogische Forschung (2014). VERA im Blick. Vertiefende

Informationen zu den Vergleichsarbeiten, zur Vorbereitung, zum fairen Vergleich, zur

Diagnosegenauigkeit und zur pädagogischen Nutzung. [Consideración de VERA. Información

complementaria de las pruebas comparativas para profundizar su preparación, comparación

justa, precisión diagnóstica y su uso pedagógico.] Recuperado de: [http://vera-web.uni-](http://vera-web.uni-landau.de/verapub/index.php?id=425)

[landau.de/verapub/index.php?id=425](http://vera-web.uni-landau.de/verapub/index.php?id=425)



Estudios e  
investigaciones

EXPERIENCIAS INTERNACIONALES SOBRE EL  
USO DE LOS RESULTADOS DE EVALUACIONES EDUCATIVAS

Es una publicación digital del Instituto Nacional para la Evaluación de la Educación.  
En su formación se emplearon las familias tipográficas Corbel y Caecilia LT Std.  
Diciembre de 2016.