

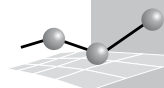
2009

"2009 / Año de Homenaje a Raúl SCALABRINI ORTIZ"

HACIA UNA CULTURA DE LA EVALUACIÓN

ONE 2009 / Censo

DiNIECE Dirección Nacional de
Información y Evaluación
de la Calidad Educativa

 **Evaluación
EDUCATIVA**

 **Ministerio de
Educación**
Presidencia de la Nación

AUTORIDADES

Presidenta de la Nación

DRA. CRISTINA FERNÁNDEZ DE KIRCHNER

Ministro de Educación

PROF. JUAN CARLOS TEDESCO

Secretario de Educación

PROF. ALBERTO SILEONI

Subsecretario de
Planeamiento Educativo

LIC. OSVALDO DEVRIES

Director Nacional de Información y
Evaluación de la Calidad Educativa

Lic. EDUARDO ARAGUNDI

Elaboración:
Área Evaluación de la Calidad Educativa

Coordinadora:
Sonia Hirschberg

Equipo responsable:
Mirta Leon
Patricia Scorzo
Jorge Novello

Diseño Gráfico:
Karina Actis
Juan Pablo Rodríguez
Coralía Vignau

ÍNDICE

Presentación	Pág. 5
--------------------	--------

Primera Parte.

Hacia una cultura de la Evaluación Educativa

Evaluación: nuevos significados para una práctica compleja, por Alicia Bertoni, Margarita Poggi y Marta Teobaldo.....	Pág. 7
--	--------

Evaluación educativa: una aproximación conceptual, por Nydia Elola y Lilia Toranzos	Pág. 15
--	---------

La evaluación educativa, un criterio consolidado, por IIPE –Buenos Aires-	Pág. 17
---	---------

Evaluación Nacional y Evaluaciones Internacionales, por el equipo de la DiNIECE	Pág. 21
---	---------

¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación?, por Alejandro Tiana Ferrer	Pág. 27
---	---------

Las evaluaciones educativas que América Latina necesita, por P. Ravela, P. Arregui, G. Valverde, R. Wolfe, G. Ferrer, F.M. Rizo, M. Aylwin y L. Wolff	Pág. 35
--	---------

Una mirada técnico-pedagógica acerca de las evaluaciones de calidad educativa, por Felipe Martínez Rizo	Pág. 49
--	---------

Segunda Parte.

El Operativo Nacional de Evaluación 2009

Marco de Referencia, por Alicia Cayssials.....	Pág. 53
--	---------

Evaluación, enfoques metodológicos, por Mariela Leones y Jorge Fasce	Pág. 57
--	---------

ADENDA

1 / Capacidades cognitivas ONE 2009	Pág. 69
2 / Evaluación Nacional y Evaluaciones Internacionales	Pág. 73
Glosario	Pág. 77



*Dirección Nacional de Información
y Evaluación de la Calidad Educativa
Ministerio de Educación*

Paraguay 1657 2do piso Of. 201
Ciudad. de Buenos Aires - C1062ACA

Tel. (+5411) 4129-1448/9 Fax. 4129-1447

E-mail: diniece@me.gov.ar

Página Web: <http://me.gov.ar/diniece>

PRESENTACIÓN

En los últimos años, el proceso de elaboración de los instrumentos de evaluación y de análisis de resultados, tanto a nivel nacional como internacional, se ha ido complejizando. Con el reemplazo de los modelos tradicionales de pruebas por el de modelos fundados en la llamada teoría de respuesta al ítem, se ha iniciado un proceso de innovación que implica la necesidad de desarrollar nuevas capacidades y tecnologías para la elaboración de instrumentos y la producción e interpretación de resultados.

Por otro lado también se advierte el desarrollo de una tendencia a nivel internacional que señala el involucramiento de mayor cantidad de actores, tanto en el diseño de los procesos evaluativos como en el uso de la información que los mismos proporcionan. Esto implica que la evaluación está dejando de ser un dispositivo que utilizan sólo los gobiernos y los investigadores para pasar a ser una herramienta para los propios actores del sistema educativo.

Es en este sentido que la DINIECE se compromete a desarrollar todo tipo de acciones para propiciar la participación y el compromiso de supervisores, directivos, maestros y familias en el camino hacia la construcción de una cultura de la evaluación en nuestro país. Se propone, asimismo, fortalecer la dimensión política de la evaluación lo que implica lograr que los resultados impacten en las prácticas educativas a través de nuevas estrategias de llegada a las escuelas y las familias y la construcción de mejores escenarios para la devolución y difusión de los resultados.

Esta publicación intenta cumplir con los objetivos arriba mencionados, por un lado su amplia difusión a funcionarios, supervisores y directores de escuelas medias, pretende contribuir a enriquecer los marcos teóricos y metodológicos, por otro presenta los debates más significativos que se están dando en el ámbito internacional.

El primer trabajo de Bertoni, Poggi y Teobaldo nos señalan la importancia de la participación de los actores en todo proceso de evaluación. Tiana nos presenta las preguntas que tenemos que considerar al iniciar un proceso de evaluación y el equipo de PREAL define las características que deben tener las evaluaciones en América Latina. Martínez Rizo realiza un análisis sobre los aspectos técnico- pedagógicos de las evaluaciones. Por último Fasce y Leones exponen los enfoques metodológicos relacionados con el censo en el último año de la escuela secundaria que se realizará en nuestro país en el mes de septiembre de este año.

En una adenda incorporamos algunos documentos elaborados por los equipos técnicos de evaluación, que consideramos de gran utilidad para el trabajo de este año: el listado de las capacidades cognitivas a considerar en el censo 2009, un análisis de las evaluaciones nacionales e internacionales realizadas en nuestro país y un glosario con los términos técnicos que utilizamos cotidianamente en nuestra tarea.

Esperamos que este material sea de utilidad para compartir en cada uno de los lugares de trabajo y para actualizar y enriquecer la tarea de cada uno de nosotros.

Primera Parte.

Hacia una cultura de la Evaluación Educativa

EVALUACIÓN: NUEVOS SIGNIFICADOS PARA UNA PRÁCTICA COMPLEJA

Los significados de la evaluación educativa: alternativas teóricas

La evaluación educativa nos confronta con el abordaje de un espacio de conflicto, que permite analizar articulaciones o fracturas entre supuestos teóricos y prácticas pedagógicas. En efecto, todo proceso de evaluación pone en evidencia múltiples aspectos, relacionados con las características y procesos de la institución educativa, los proyectos institucionales, los estilos de gestión, las propuestas curriculares y editoriales, las particularidades de los docentes y de los alumnos, etc. Podemos afirmar, sin lugar a dudas, que nos enfrentamos a un punto neurálgico de la relación educativa.

Más allá de las diferentes perspectivas que han abordado la cuestión de la evaluación, nos parece importante señalar que ésta constituye siempre una actividad de comunicación en la medida en que implica producir un conocimiento y transmitirlo, es decir, ponerlo en circulación entre diversos actores involucrados. En tanto acción comunicativa, la evaluación padece las mismas vicisitudes que toda comunicación entre los sujetos sociales implicados en ella.

Aun cuando pueda plantearse en diferentes niveles, como el del sistema educativo, la institución o el aula, la evaluación implica siempre una serie de aspectos comunes que siguen una secuencia lógica y cronológica aunque, en la práctica, algunos de ellos sean omitidos. Nos referimos a:

- el relevamiento de información a través de diversos procedimientos como la observación, los documentos, las producciones, etcétera;
- el análisis de los datos según marcos de referencia que orientan la "lectura" de éstos;
- la producción de conclusiones, en algunos casos expresadas en juicios de valor, que traduce el análisis en proposiciones sobre el "objeto" evaluado, o, en otros casos, la producción de datos cuantitativos, "cifrados", que muestran algún otro aspecto de dicho objeto;
- la comunicación a los actores involucrados en el proceso evaluativo o la divulgación a otros de las conclusiones elaboradas;
- aunque no necesariamente se deriva de los puntos mencionados, la toma de decisiones para intervenir activa e intencionalmente en los procesos y resultados sobre los cuales se emitió el juicio evaluativo.

A pesar de que las instancias mencionadas constituyen una secuencia, su ejecución no es, necesariamente, lineal. El análisis de los datos, por ejemplo, puede requerir el relevamiento de alguna información adicional o complementaria; la comunicación de los resultados a los actores implicados puede llevar a cierta reconsideración de las conclusiones, etcétera.

Por lo demás, según cuál sea el paradigma que da sentido a la evaluación educativa, algunos puntos se priorizarán sobre los otros o adquirirán diferentes significados.

La articulación del campo educativo con otros campos: implicancias para la evaluación

Una definición amplia de evaluación pone el acento en que esta actividad consiste en la atribución de un juicio de valor a una realidad observada.

Ahora bien, cuando la realidad que se evalúa se vincula con los aprendizajes de los alumnos en el contexto de la vida escolar, el juicio de evaluación que se emite compromete a otros niveles con los que se articula el ámbito escolar. Esto es así porque, como veremos, los ámbitos del aula y de la escuela se inscriben en el campo educativo, el cual, a la vez, se articula con otros campos sociales y se encuentra, en distintas situaciones históricas, sobredeterminado por alguno de ellos.

Empleamos la noción de "campo" en el sentido en que la define Pierre Bourdieu, es decir, como "espacios de juego históricamente constituidos con sus instituciones específicas y sus leyes de funcionamiento propias". En este sentido, se habla de campo educativo, de campo económico, de campo político, etcétera.

Volviendo entonces al campo educativo, podemos decir que se articula con otros campos, como el científico, el artístico o el tecnológico, y que se encuentra, en nuestro contexto histórico-social, sobredeterminado, en buena medida, por el campo político y por el campo económico.

Si dentro del campo educativo nos situamos ahora en el ámbito del aula, podremos observar que lo que sucede dentro de ésta tiene repercusiones extraescolares, que se explican por las articulaciones y sobredeterminaciones antes mencionadas.

Este fenómeno es evidente en el caso de la evaluación, no sólo por la articulación entre el campo educativo y otros campos, sino también por la función social que aquella cumple. Efectivamente, en la medida en que la evaluación sanciona la adquisición o no de los aprendizajes que se

consideran necesarios para el desempeño social, sus resultados repercuten no sólo sobre el sujeto evaluado y su entorno inmediato en las circunstancias presentes, sino que tienen un efecto "expansivo", no necesariamente visible, sobre su inserción futura en otros campos.

Podemos ver cómo tiene lugar este proceso situando el análisis en el nivel del aula.

La observación de lo que sucede con la evaluación de los aprendizajes en el aula permite a un observador atento construir un entramado de relaciones entre actores y entre campos, que se constituyen con la puesta en marcha del dispositivo de evaluar y, sobre todo, con los resultados derivados del mismo. En este sentido, el ámbito del aula involucra no sólo a los protagonistas que en buena medida lo conforman, a saber, los docentes y los alumnos, sino también a otros actores sociales, cuya visibilidad depende del grado de proximidad que guarden con los miembros de la relación educativa: padres, familiares, amigos, futuros empleadores, etcétera.

La evaluación, en su nivel más visible, opera como un "eje" que articula el "juego" entre tres tipos de actores: docentes, alumnos y padres, o quienes desempeñen el rol sustituto de éstos. Pone en contacto a "los protagonistas" de los actos educativos con los padres, que son actores "puente" entre lo escolar y lo extraescolar. En realidad, la evaluación es la instancia por excelencia, en términos de los "asuntos" que suelen poner en relación a los tres tipos de actores, porque sus resultados tienen repercusión sobre la trayectoria educativa de los alumnos, sobre sus posibilidades futuras más allá del campo educativo, sobre las expectativas de los padres y sobre la responsabilidad del maestro¹.

Por esta razón, docentes, alumnos y padres entran en relación a propósito de las calificaciones. En la escuela "juegan" el "juego" de las notas porque en las instituciones educativas se producen, negocian y distribuyen calificaciones, que circulan luego en otros campos, bajo la forma de certificaciones, con valor en el mercado.

Estos procesos de producción, negociación y distribución de las calificaciones asumen características particulares en el interior de las escuelas y de las aulas. Las "reglas de juego", formales o informales, explícitas o implícitas, que se configuran entre docentes, alumnos y padres a propósito de la evaluación, se relacionan con las características institucionales y con la manera en que aquellos actores desempeñan sus roles, es decir, con la manera en que "juegan el juego" en tanto sujetos vinculados por la evaluación. En esto inciden las concepciones o "ideas" que tienen acerca de ella, la importancia que le asignan, las estrategias de sobrevivencia desarrolladas para jugar el juego en las condiciones dadas, etcétera. Así, por ejemplo, los docentes califican a sus alumnos en función de sus concepciones sobre la evaluación, sus expectativas sobre los alumnos, su mayor o menor contaminación con las rutinas escolares. Los alumnos, por su parte, responden a las evaluaciones según el grado de interés que tengan por lo aprendido, según las estrategias que hayan desarrollado para "pasar" una prueba o un examen, etc. Por otro lado, los padres se interesan por las calificaciones de sus hijos y ejercen, según los casos, presiones sobre éstos o sobre los maestros.

En un nivel menos visible, la evaluación pone en evidencia la vinculación entre los dos actores presentes en la relación pedagógica (docentes y alumnos) y un actor virtual: los futuros empleadores. Esta relación se percibe con mayor dificultad cuanto más alejado está el alumno del mercado laboral. Sin embargo, en la relación pedagógica siempre está en juego el valor social de los alumnos en tanto futuros trabajadores y ello remite, necesariamente, a la relación entre lo que se enseña y aprende en la escuela, por un lado, y las exigencias que se derivan del mundo de la producción, por otro. En este sentido, "juegan el juego" los docentes, los alumnos y un actor social implícito, constituido por los agentes del mercado del empleo.

Más allá de las condiciones económico-sociales que inciden en la distribución desigual del capital cultural, los procesos que "fabrican" el fracaso escolar en el interior de las escuelas (Ph. Perrenoud, 1990) demoran o interrumpen, por repetición o deserción, el juego de los alumnos en tanto actores de ese proceso de formación y, por ello, ponen en peligro su valor social en el futuro mercado de trabajo. Esto sucede, básicamente, por dos razones: en primer lugar, porque el capital cultural en estado incorporado es insuficiente, es decir porque los habitus como disposiciones duraderas relacionadas con conocimientos específicos, valores, habilidades, etc., han carecido del tiempo necesario para su adquisición y, en segundo término, porque el capital cultural en estado institucionalizado, bajo la forma de certificaciones o títulos, no ha llegado a obtenerse².

En cuanto a las relaciones entre los actores de este juego, son particularmente notorias en nuestros días: los empresarios juzgan el valor y la calidad de los contenidos educativos y de la formación de los docentes; éstos, a la vez, discuten los parámetros en los que aquéllos fundamentan sus juicios y los alumnos, a su turno, se incorporan a este diálogo cuando están en condiciones de comprender la importancia que tiene para su futuro la calidad de la enseñanza a la que tienen acceso, y esperan o requieren transformaciones sustantivas de ella.

Por último, en el entramado de relaciones al que venimos haciendo referencia, la evaluación pone en evidencia que, en el orden social, los docentes son trabajadores que deben responder a las orientaciones que formulan los responsables de la toma de decisiones en nombre de las necesidades públicas. Lo que está en juego en este orden es, teóricamente, la satisfacción de necesidades sociales específicas; lo cual implica, en el campo educativo, retraducciones de estas últimas en diversos niveles de especificidad y en múltiples dimensiones. Ello significaría, por ejemplo, que los fines y objetivos del sistema educativo sean congruentes con las necesidades y demandas sociales relevadas; que exista

¹ Bourdieu P.: *Cosas Dichas*, Barcelona, Gedisa, 1988

² Bourdieu P., "Les trois états du capital culturel", *Actes de la Recherche en Sciences sociales*, París N° 30, noviembre de 1979.

coherencia entre las finalidades y objetivos formulados y la estructura del sistema educativo y sus formas de gestión; que la propuesta curricular sea compatible con los aspectos mencionados y significativa en términos de la demanda y requerimiento sociales, etcétera.

La identificación de los actores del ámbito escolar y el análisis de la imbricación de sus prácticas con las de otros actores que operan en otros campos muestran por qué la evaluación remite necesariamente a otros espacios extraescolares como el campo económico y el campo social. En este sentido, puede decirse que la evaluación está sobredeterminada y es multidimensional.

Estas características se pueden advertir cuando se analizan algunos de los procesos que se inician en el ámbito pedagógico. Se observa, por ejemplo, que, cuando los docentes evalúan las realizaciones de sus alumnos atribuyéndoles calificaciones, van creando, en el proceso de asignación de puntajes, realidades inexistentes hasta ese momento: "buenos alumnos", "malos alumnos", etc. De esta manera, no sólo se atribuye un valor a lo que los alumnos producen, sino a los alumnos mismos, con lo que se cae en el conocido riesgo de anticipar el destino escolar del sujeto y con ello su futuro valor social (en el espacio económico, particularmente).

Por otra parte, aun cuando los docentes no vinculen directamente estas realidades a su propia actividad, ésta resulta de alguna manera evaluada; particularmente los malos o insuficientes resultados que alcanzan sus alumnos suelen llevarlos a cuestionar la pertinencia social de su tarea en el campo social y su rentabilidad en el campo económico.

De esta forma, cada acto de evaluación, por las sobredeterminaciones que operan sobre él, "se encuentra, como las muñecas rusas, en el corazón de una serie de otros actos que lo envuelven"³.

Se podría proceder a un análisis parecido con los alumnos y con los padres si se quisiera analizar las implicaciones económicas y sociales que tienen las evaluaciones para estos actores.

Los usos sociales de la evaluación

La multidimensionalidad de los actos evaluativos, producto de sus articulaciones y sobredeterminaciones, conduce a poner en duda la unidad y coherencia de lo que se designa con el término "evaluación". Algo similar ocurre cuando se analizan los diversos usos sociales que se hacen de ésta.

Cuando se considera el problema de la utilización de los resultados, puede observarse que el empleo de ellos está vinculado básicamente a:

- las intenciones de la evaluación;
- las acciones derivadas de los resultados.

■ Las intenciones de la evaluación

Si nos preguntamos por qué se evalúa, la cuestión remite inmediatamente a la intención de quien construye el objeto de evaluación (conocer las respuestas de los alumnos a un conjunto de situaciones problemáticas, por ejemplo).

Si, en cambio, el interrogante está referido a para qué se evalúa, la preocupación refiere a los efectos de la evaluación sobre la acción (qué se hará con los resultados).

En el primer caso, los propósitos o intenciones de evaluar dependen de cuál sea su concepción sobre la evaluación.

Las intenciones más reconocidas son las de medir, apreciar y comprender, que se corresponden con sus respectivas "filosofías" subyacentes.

La intención de medir el objeto está presente cuando éste se considera desde la perspectiva del desempeño o actuación (performance) del sujeto. Pero resulta imposible "medir" con "objetividad" las realizaciones de los alumnos en una prueba, cuando se pretende dar cuenta de sus conocimientos y competencias, porque los comportamientos educativos no pueden ser aprehendidos de manera indiscutible con un instrumento de medición. No hay relación, en este caso, entre las propiedades del instrumento que se quiere emplear y las propiedades del objeto que se pretende medir.

Jacques Ardoino y Guy Berger⁴ establecen una distinción entre la evaluación estimativa, que constituye un enfoque que prioriza lo cuantitativo, y la evaluación apreciativa, que privilegia lo cualitativo.

En el primer caso, la intención es efectuar una lectura de lo real lo más próxima posible a la medición. Como no se posee una unidad de medida indiscutible, se trata, en primer término, de circunscribir, de delimitar ciertas características de la realidad a estudiar y de estimarlas luego de la manera más objetiva posible valiéndose de consideraciones de tipo cuantitativo. Es conocida, en este sentido, la experimentación con un nuevo programa de estudios o con nuevos métodos de enseñanza para estimar en qué medida mejoran, a partir de su aplicación, los resultados educativos alcanzados hasta el momento.

También se evalúa para apreciar, para determinar el valor de un objeto o realidad dada, con referencia a ciertos criterios preexistentes. La "evaluación apreciativa" puede dar lugar, sin embargo, al desarrollo de dos orientaciones que se corresponden con dos "filosofías" diferentes. Por una parte, a la evaluación apreciativa con un modelo predeterminado, que presupone un referente previo, anterior a toda recolección de información. El referente orienta la lectura de la realidad, y la evaluación consiste en la búsqueda de los indicadores que permitan dar cuenta del estado del objeto con relación al referen-

■ 3 Hadji C., *L'évaluation des actions éducatives*, Paris, PUF, 1992, p. 22.

■ 4 Ardoino J. y Berger G., *D'une évaluation en miettes à une évaluation en actes*, Paris, ANDSHA-Matrice, 1989.

te. En la evaluación de los alumnos, éste puede consistir, por ejemplo, en los perfiles de logro como perfil tipo previamente establecido. La evaluación procuraría apreciar en qué medida las realizaciones de los alumnos se vinculan a los indicadores seleccionados como indicadores de logro. Cuando la evaluación parte de criterios predeterminados, responde, en general, a las características descriptas y sigue procedimientos análogos.

Por otro lado, la evaluación apreciativa sin modelo predeterminado expresa una concepción de la evaluación que se fundamenta en la interpretación. Es aquella que se interroga por el sentido. Evaluar el funcionamiento de una clase o de una escuela implica, dentro de esta metodología, construir -en el proceso mismo de investigación- el referente apropiado, es decir, aquel que permita aprehender la singularidad del aula o de la escuela que se evalúan. Se apunta a comprender el objeto, no a juzgarlo. Se evalúa, entonces, para volver inteligible la realidad, para aprehender su significación.

En este contexto, evaluar las realizaciones de los alumnos significa comprender sus maneras de resolver las situaciones planteadas, considerar sus errores como indicadores de problemas a descifrar, etc. En el mismo sentido, más que evaluar el cumplimiento de los objetivos, interesa comprender las razones por las cuales éstos son o no alcanzados y entender qué representa esta situación en términos didácticos.

■ Las acciones derivadas de los resultados

Los resultados de la evaluación tienen relación directa con el objetivo de ella, con la pregunta de para qué se evalúa.

En este sentido, es importante tener en cuenta que la relación entre quién o quiénes evalúan y el objeto de evaluación (los saberes de los alumnos, por ejemplo) no es independiente del contexto en el cual aquélla se realiza.

En efecto: cualquiera sea la forma que adopte, la evaluación se inscribe siempre en un ámbito de decisiones. Aun en sus formas más frecuentes y simples, como en el caso de la evaluación implícita, que tiene lugar en la interacción cotidiana entre maestros y alumnos, aquélla sirve al docente para tomar decisiones sobre sus prácticas en el aula. Tenga o no conciencia de que la realiza, el maestro releva una información a partir de la cual organiza la interacción con sus alumnos en términos de lo que considera mejores condiciones para el aprendizaje.

Cuando la evaluación tiene carácter institucional, en cambio, su modalidad es explícita; es un acto deliberado, organizado, que se efectúa empleando metodologías e instrumentos de carácter variado, generalmente complejos. En este caso, el uso social externo o interno de los resultados dependerá, en parte, del tipo de información que permitan obtener los instrumentos empleados y, en parte, de las decisiones previamente adoptadas, o bien de aquellas que se adopten a partir de la obtención de los datos. Puede ocurrir también que el uso que se haga de la información obtenida sea producto de una transacción entre las decisiones iniciales y las que se derivan

del conocimiento de los datos, lo que puede conducir, por ejemplo, a suministrar informaciones parciales.

La evaluación está siempre relacionada, entonces, con un contexto decisional dado. También el uso interno de los resultados puede ser diverso según las finalidades perseguidas por quien evalúa y el tipo de evaluación empleada.

Cuando se trata de una evaluación **diagnóstica**, el docente procura información acerca de los saberes y competencias que poseen sus alumnos en términos de requerimientos necesarios para una secuencia futura de aprendizajes. En este caso, los resultados obtenidos constituyen una información de base para adoptar las decisiones que se estimen más adecuadas: reforzar los contenidos o proseguir con la secuencia iniciada.

Cuando se emplea la evaluación **sumativa** para comprobar en qué medida los alumnos han adquirido los conocimientos esperados y las competencias correspondientes, la información orienta la decisión de promoverlos de grado, por ejemplo, o de otorgarles una certificación que acredite el término del nivel.

En la evaluación **formativa**, la decisión está directamente vinculada a la selección y puesta en práctica de secuencias de contenidos y de estrategias pedagógicas que se consideran como las más adecuadas para mejorar los resultados obtenidos. En este sentido, la información que provee la evaluación sirve para fundar decisiones pedagógicas.

En resumen: podemos decir que tanto el **"por qué" se evalúa** (intenciones) como el **"para qué" se evalúa** (uso de los resultados) determinan el tipo de evaluación empleada. Por otra parte, según la forma que adopte la evaluación, los resultados permitirán tomar decisiones de orden estrictamente pedagógico (como en el caso de la evaluación diagnóstica o formativa) o decisiones vinculadas a la certificación, la acreditación o a políticas educativas orientadas hacia el sistema (como en el caso de la evaluación sumativa).

Desde la perspectiva de los docentes, estas últimas formas de evaluación tienen, obviamente, menos interés que aquellas que aportan información directa para su tarea. En el mismo sentido, interesa menos la "objetividad" de los instrumentos de evaluación que su **utilidad**, en términos de la información que permiten obtener sobre las producciones de los alumnos.

Algunas observaciones sobre la acción de evaluar

Conforme a lo que se ha venido señalando, podría decirse que evaluar implica, siempre, tomar distancia de la realidad que se analiza para poder pronunciarse sobre ella en un contexto decisional dominante (Hadji, 1992).

La afirmación "poder pronunciarse acerca de la realidad" requiere ciertas precisiones.

La evaluación como lectura orientada: la construcción del referente

En primer lugar, evaluar supone efectuar una **lectura orientada** sobre el objeto que se evalúa, en función de la cual el evaluador se "pronuncia sobre la realidad". Dicho de otra manera, no existe una lectura directa de la experiencia. Hay siempre un proceso de interacción entre el evaluador y la realidad a evaluar. En ese proceso, en función de las propiedades de la realidad a evaluar, el evaluador construye el referente, es decir, aquello con relación a lo cual se va a efectuar la evaluación, aquello que le permitirá "pronunciarse sobre la realidad" que evalúa.

En la investigación a la que haremos referencia en el transcurso de este libro, construimos el referente a partir de los datos relevantes mediante la administración de una encuesta a docentes, en la que indagamos, entre otros aspectos, cuáles eran los contenidos que realmente se enseñaban. Pudimos así construir el referente, que se configuró como **currículum real**. Más adelante explicaremos cómo tuvo lugar ese proceso.

Es necesario precisar ahora que en la construcción del referente intervienen distintos factores. Por una parte, inciden las expectativas que conciernen al propio objeto evaluado como, por ejemplo, el nivel de rendimiento esperado en los alumnos. En éste sentido, cuando se evalúa una realidad dada, se procura comprobar si ésta presenta las características que se esperan de ella. Por otra parte, la constitución del referente se vincula a la concepción de evaluación que se sustenta, los propósitos que se derivan de ésta y el contexto decisional en que se inserta el proceso evaluativo. Veamos este punto.

Si se trata, por ejemplo, de la construcción de un sistema de evaluación en una jurisdicción dada, el referente último puede estar constituido por el **currículum prescripto**. Sin embargo, por razones que obedecen estrictamente a la concepción de evaluación subyacente, a los propósitos derivados de ella y al contexto decisional en el que se inserta la evaluación, se puede -en una primera etapa- partir de un referente constituido por aquello que los maestros afirman enseñar en los distintos grados del nivel que se evalúa. En este caso, el relevamiento del **mapa curricular real** permite construir el referente y especificar los criterios para evaluar las realizaciones de los alumnos.

En una segunda etapa, el referente puede ser lo que debería enseñarse según el juicio de los maestros (recogido mediante instrumentos adecuados) y según los resultados que arroje la evaluación de lo realmente enseñado.

En esta metodología de evaluación, el referente tiene un carácter provisorio y se traslada progresivamente del campo "del ser" (lo que se enseña) al del "deber ser" (lo que debería enseñarse).

El carácter progresivo en la construcción de los referentes permite instancias de evaluación que, partiendo de lo real, incluyen, progresivamente, lo prescripto. Estos principios metodológicos responden a una concepción orientada a recabar, desde el inicio, información próxima al campo de experiencia de los protagonistas

de la relación pedagógica para que los resultados obtenidos a través de la evaluación permitan a los docentes reorientar los procesos de enseñanza.

Por otra parte, una evaluación que opera por niveles como los señalados facilita el análisis de las sucesivas transposiciones del currículum prescripto, en las expectativas de los maestros sobre lo que debería enseñarse, y en el currículum real.

De esta forma, es posible introducir, en las metodologías de evaluación, modificaciones que tiendan a poner los procesos evaluativos al servicio de la acción pedagógica. Sólo con la construcción del referente hemos dado un ejemplo acerca de cómo pueden comenzar a producirse cambios respecto a "lo que se pone en juego" en el ámbito pedagógico. Más adelante mostraremos que la metodología genera también transformaciones en los papeles de los actores y en las reglas del juego que los relacionan.

Volviendo ahora al análisis del proceso de evaluación, es necesario señalar que al referente se añade un conjunto de criterios que guían tanto la construcción de los instrumentos de evaluación como los procesos de valoración posteriores.

Por otra parte, los criterios también operan en los procesos de selección de lo que se quiere evaluar e intervienen, asimismo, en la construcción de los indicadores y de los datos resultantes de la evaluación.

En este sentido, **tanto el objeto que se evalúa como el proceso de valoración son construidos por el sujeto que evalúa.**

En la evaluación, como en cualquier otro acto de conocimiento, "no se capta nunca directa y totalmente la realidad examinada. Ésta es siempre objeto de un proceso previo de definición o de elección de los caracteres que permitirán aprehenderla" [...] "Formulado de otra forma, no basta con decir que la evaluación lleva a la realidad; todavía falta construir o adoptar los indicadores de la realidad considerada. Desde este punto de vista, la evaluación plantea el conjunto de cuestiones planteado de forma más amplia por los procesos de producción de conocimientos"⁵.

Veamos ahora, a manera de ilustración, algunos de los pasos necesarios en la construcción del objeto. Una vez que el propósito de la evaluación ha sido determinado (conocer, por ejemplo, las realizaciones de los alumnos en un área de conocimiento y en un nivel específico), se hace necesario fijar los criterios para evaluar las respuestas producidas por los alumnos y construir los indicadores correspondientes (tipo y nivel de conocimientos esperables, competencias requeridas, etc.). En este proceso, la selección de los contenidos curriculares por evaluar supone también partir de criterios que permitan efectuar un recorte técnico en relación con los contenidos enseñados y con las respuestas valoradas como deseables. Así, la decisión implica poner en juego una concepción pedagógica determinada y operar con un marco epistemológico específico.

5 Barbier J.M, *La evaluación de los procesos de formación*, Madrid, Paidós, 1993, p 66.

El proceso de construcción del objeto supone, entonces, proceder por recortes sucesivos de la realidad, los cuales se realizan conforme a criterios con independencia de que estén o no explicitados. Por esta razón, los indicadores que se construyen no cubren nunca la totalidad de la realidad que se evalúa; como se habrá comprendido, al evaluar siempre se introduce un esquema de lectura de la realidad que circunscribe, recorta, reduce lo observable.

En cuanto a los procesos de valoración de los resultados, nos remitimos a lo señalado anteriormente acerca de que la evaluación, en tanto adjudicación de un juicio de valor a una realidad dada, supone siempre una lectura orientada, "filtrada" por el referente y por los criterios que constituyen una explicitación de este último.

Es conveniente aclarar, asimismo, que los procesos de valoración dependen, por otra parte, de la pertinencia de los instrumentos de evaluación construidos, es decir, de la adecuación entre el tipo de instrumento elaborado y las características de los procesos educativos que se desee aprehender.

Como veremos más adelante, la construcción del objeto y los procesos de valoración pueden ser el resultado de una metodología como la que fragmentariamente presentamos como ejemplo, de un enfoque más cualitativo (en el cual el objeto y la valoración se construyen en el proceso de evaluación) o de una combinatoria de ambos paradigmas.

La construcción de los resultados

Con respecto al producto de la evaluación, también concierne al evaluador construir los resultados de la evaluación, construir los datos, "lo referido" (con relación al referente). Ello significa relevar en la realidad los signos -los indicadores- que den cuenta de la presencia de los criterios previamente establecidos.

Desde este punto de vista, lo que se denomina **prueba de evaluación** tiene por función hacer aparecer los signos buscados⁶.

Para que esto sea posible, las "pruebas" deben satisfacer el requisito de la pertinencia a fin de que pueda observarse la relación existente entre los indicadores (estructura del razonamiento de los alumnos, estrategias que ponen en juego para la resolución de los problemas, etc.), el referente y los criterios en función de los cuales se evalúa.

Si la metodología adoptada se fundamenta en una concepción según la cual la construcción del referente y de los resultados debe responder a la intención de contribuir al mejoramiento de las prácticas educativas, también el juicio del evaluador debe servir a ese propósito. En este sentido, más allá de la utilidad de las calificaciones obtenidas, en términos de información pertinente para la gestión y de interés para los padres, docentes y alumnos, resulta indispensable una devolución cualitativa de la información.

Como cierre de este apartado y en función de lo que se ha analizado hasta ahora sobre la multidimensionalidad de la evaluación y la variación de sus usos sociales,

quisiéramos señalar que pueden identificarse dos características que permanecen como elementos constantes en toda actividad evaluativa:

- 1. Siempre constituye una lectura orientada** (tanto cuando asume formas de control como cuando se trata de la evaluación apreciativa con referente predeterminado o de la evaluación interpretativa).
- 2. Siempre consiste en un pronunciamiento acerca de la realidad.**

Los múltiples significados de la palabra "evaluación"

Como acción comunicativa, la evaluación nos permite una aproximación a ella desde el punto de vista semántico⁷. Tanto en el diccionario, como en las acepciones más habituales del término, o aquellas asociadas con él, aparecen involucrados diferentes significados. Recuperemos algunos de ellos:

Verificar / medir / valorar / comprender / aprehender / conocer
 Juzgar / comparar / constatar / apreciar / decir / ayudar
 Cifrar / interpretar / estimar / experimentar / posicionar / expresar

No obstante, los diferentes términos -en ocasiones complementarios, en otras opuestos- implican distintas perspectivas desde el punto de vista de la evaluación educativa. Analicemos algunos de ellos, tratando de poner algún orden en la lectura.

Podríamos afirmar que de los términos enunciados pueden desprenderse, por lo menos, dos perspectivas sobre la evaluación.

En primer lugar, evaluar parece involucrar el hecho de medir con precisión; en este sentido, expresar una cantidad precisa, cifrada. En otras palabras, alude a expresar una medida cuantificada.

En segundo lugar, evaluar implica operaciones como estimar, apreciar, aprehender, en otros términos, pronunciar un juicio cualitativo y, eventualmente, aproximativo sobre una realidad.

"Evaluar es en consecuencia un término bien singular que puede expresar una cosa y su contrario: lo preciso y lo aproximado, lo cuantitativo y lo cualitativo"⁸. Es imposible, por lo tanto, no destacar el grado de ambigüedad presente en la evaluación y lo que ello implica a la hora de generar y producir prácticas evaluativas en las instituciones escolares.

⁶ Hadji, *op. cit.* p. 36.

⁷ Barlow M., *L'évaluation scolaire. Décoder son langage*, Lyon, *Chronique Sociale*, 1992, p. 60 y s.s.

⁸ Barlow M., *op. cit.*, 1992, p. 63.

Por otra parte, más allá de estas dos grandes posiciones señaladas, aparecería, en algunos de los términos que se asocian a la evaluación, la idea de ayuda, sostén o apoyo, con lo que ello implica desde el punto de vista educativo.

La evaluación y el control

En función de esta ambigüedad del término, que refleja lo que sucede en las prácticas evaluativas, Jacques Ardoino y Guy Berger presentan la evaluación como un "Jano moderno" con un doble perfil: un lado muestra un perfil filosófico, en la medida en que toda evaluación plantea el problema del valor, del sentido y de la significación de aquello que se evalúa (lo que requiere un tratamiento cualitativo); el otro lado muestra un perfil técnico, ya que la evaluación constituye un dispositivo compuesto por métodos, técnicas e instrumentos empleados "para dar cuenta y rendir cuenta" (de manera cuantitativa) de los resultados obtenidos. Ese perfil se asocia con la noción de control.

Para los autores mencionados, sin embargo, ambas nociones pertenecen a dos órdenes diferentes y a dos paradigmas epistemológicos distintos. Veamos la significación de cada uno de los términos.

La palabra "control" tiene su origen en la lengua francesa, en la contracción de la expresión "contrerôle", empleada para designar el "doble registro" necesario para autenticar las escrituras. Adoptada por numerosas lenguas, la expresión ha conservado a lo largo del tiempo el sentido administrativo de su origen.

En nuestra vida cotidiana la palabra expresa múltiples y diversas situaciones: control administrativo, control financiero, control fiscal, control policial, control aduanero, control industrial, control de calidad, etcétera.

La mayor parte de los instrumentos de medición que empleamos son instrumentos de control. También implican operaciones de control los mecanismos cibernéticos que regulan la transmisión y recepción de información en sistemas simples o complejos.

En el campo educativo, los exámenes, el registro de asistencias, las amonestaciones, etc., son procedimientos explícitos de control.

Esta diversidad de usos y acepciones ha conducido a considerar el control —como un sistema, un dispositivo y una metodología, constituidas por un conjunto de procedimientos que tiene por objeto (y objetivo) establecer la conformidad (o la no conformidad), y aun la identidad, entre una norma, un patrón, un modelo y los fenómenos u objetos con los que se los compara, y en ausencia de esta conformidad o identidad, establecer la medida de su diferencia"⁹.

Es importante retener, para la comprensión de este término y su ulterior comparación con el de evaluación, que el control se efectúa a partir de un elemento externo y anterior (desde el punto de vista lógico, no necesariamente cronológico) al acto mismo del control. Los exámenes y las amonestaciones, por ejemplo, existen, en tanto instrumentos, con anterioridad a la instancia de

su aplicación y con independencia de ella.

De manera totalmente opuesta, la evaluación -de acuerdo con su propia etimología- implica la problematización sobre los valores y sobre el sentido de lo que ocurre en la situación observada. Evaluar es aprehender las significaciones propias, particulares de los actos humanos. Importa más, en esta acepción, la aprehensión de los significados que la coherencia o conformidad con un modelo dado.

La cuestión del sentido introduce otra diferencia sustancial: mientras el control siempre opera a partir de un solo y único referente, que es el patrón de medida como norma homogeneizadora de lo que se mide, la evaluación es multireferencial en tanto debe aprehender significaciones heterogéneas. Está siempre abierta al sentido y, por lo mismo, es inacabada.

A diferencia de los acontecimientos que se miden a través de instrumentos de control, la evaluación se aplica a procesos humanos en los cuales la dimensión temporal es histórica e irreversible porque consiste en el tiempo vivido. Por lo tanto, no puede estar referida a estructuras objetivas en el mismo sentido que el control, que mide los acontecimientos conforme al tiempo cronológico: un año, un mes, una semana, etcétera.

La comprensión de las características específicas de los procesos de control y evaluación permite advertir que la oposición entre ambos es, en realidad, una cuestión que excede el problema de los métodos y de las técnicas de cada uno. Se trata, más bien, de un problema epistemológico que se traduce en la confrontación entre dos paradigmas distintos.

Sin embargo, más allá de la oposición entre ambos procesos, se ha observado que ellos comparten ciertas funciones comunes, indispensables para la regulación crítica de la acción. Regulación, en el sentido que permiten ajustar las acciones con relación a un objetivo establecido, y crítica, porque dicho ajuste se realiza a partir de una lectura orientada.

Esta propiedad compartida por ambos procesos conduce a considerar que la regulación crítica de la acción puede ser concebida a lo largo de un "continuum" constituido por dos polos: el control, por un lado, y la evaluación interpretativa, por el otro.

Dentro de ese continuum se ubica, hacia el polo del control, la evaluación estimativa, que, como se vio anteriormente, se inclina hacia lo cuantitativo. En cierta forma, también la evaluación apreciativa con un referente predeterminado tiene ciertas propiedades que la relacionan con el control en el sentido de que, si el referente está totalmente explicitado, el proceso evaluativo consiste en verificar, en la realidad, en qué medida aquél ha sido alcanzado.

En cambio, la evaluación apreciativa sin referente predeterminado se ubica en el polo de la evaluación, toda vez que carece de un modelo referencial y apela a la interpretación de la realidad observada. Se ubican en

⁹ Ardoino y Berger, *op. cit.*, p. 12.

esta línea, fundamentalmente, las orientaciones basadas en el paradigma interpretativo, que, como veremos más adelante, plantean la construcción del referente en el proceso mismo de evaluación.

La introducción del concepto de regulación crítica permite separar la evaluación estimativa de la medición en sentido estricto. Entre ambas existe una frontera, ya que el objetivo de toda medición es hacer una descripción cuantitativa de la realidad, efectuar una lectura de ella lo más "objetiva" posible. La evaluación estimativa, en cambio, no tiene como único objetivo medir, sino emitir un juicio a partir de los datos que la medición provee. Este tipo de evaluación pone de manifiesto que la realidad que se evalúa no puede -en sentido estricto- ser medida, pero, además, que la intención dominante no es sólo conocer sino regular. Recordemos el ejemplo que dimos a propósito de este tipo de evaluación cuando se empleaba en situaciones experimentales a partir de la incorporación de un nuevo programa o método, destinado a mejorar los aprendizajes. En ese caso, la información suministrada por la evaluación para estimar la medida en que la innovación curricular o metodológica mejoraba los resultados estaba al servicio de la regulación del proceso de enseñanza, permitiendo así introducir modificaciones en este último.

A pesar de que tanto el control como la evaluación cumplen una función de regulación crítica para la acción, ambos procesos pueden ser necesarios porque dicha función tiene un carácter distinto en cada caso.

Habitualmente, sin embargo, suelen utilizarse los conceptos de medición y evaluación de manera confusa y poco discriminada. El término "evaluación", por ejemplo, es empleado de manera genérica y subsume prácticas muy heterogéneas, que incluyen las de medición en el sentido estricto de control. Parece adecuado suponer, como lo hacen Ardoino y Berger, que las connotaciones negativas que afectan al proceso de control inhiben la explicitación de su uso en las situaciones en que se lo

emplea, con lo cual se induce a errores o confusiones conceptuales, por una parte, y se anula o limita la posibilidad de su uso, por otra¹⁰.

En resumen, el control y la evaluación son funciones interdependientes, idealmente complementarias pero, sin embargo, teóricamente distintas por los paradigmas en que se fundan. En efecto, mientras en los casos en que se opera a partir del control se procede a establecer el grado de conformidad de un objeto o de una situación con relación a una norma o a un sentido dado, en la evaluación, el sentido se construye en el proceso de interacción y de intercambio de significaciones.

Ambos procesos son necesarios para la inteligibilidad y la regulación de las acciones, aunque no podrían ser ubicados en el mismo registro, porque el tipo de información que proveen es diferente.

La consideración de algunas de las características propias de las metodologías del control y de la evaluación muestra sus diferencias y su posible complementariedad.

En el primer caso, el análisis de situaciones y de comportamientos, individuales o colectivos, supone interesarse por indicadores, contruidos en función de criterios que, a la vez, remiten a los paradigmas fundantes. A este conjunto se agregan los instrumentos para medir o apreciar los efectos o resultados.

Desde otra perspectiva, que se relaciona con los análisis de tipo cualitativo, interesa describir e interpretar el sentido de las situaciones observadas.

La complementariedad de los enfoques permitiría indagar sobre las causas de los resultados cuantitativos obtenidos, es decir, sobre el fundamento de ellos, incorporando el uso de instrumentos que permitan recoger información adicional de carácter cualitativo. Tal complementariedad permitiría trabajar en el nivel de lo implícito, de los presupuestos, de los esquemas de valoración, etc., que operan en la dinámica de la interacción de los actores, en las distintas fases del proceso evaluativo.

■ 10 Ardoino y Berger, *op. cit.*

“EVALUACIÓN EDUCATIVA: UNA APROXIMACIÓN CONCEPTUAL”

Resulta oportuna, en términos generales, la definición propuesta por T. Tenbrink: Evaluación es el proceso de obtener información y usarla para formar juicios que a su vez se utilizarán en la toma de decisiones.

a) Las funciones de la evaluación

En términos generales se pueden reconocer diferentes funciones frecuentemente atribuidas a la evaluación, las mismas no son excluyentes sino complementarias y algunas se explican a través de las ideas más generalizadas que se tienen sobre la evaluación y otras se relacionan directamente con un concepto más completo y complejo de estos procesos.

- **función simbólica:** los procesos de evaluación transmiten la idea de finalización de una etapa o ciclo; se asocia con frecuencia la evaluación con la conclusión de un proceso, aún cuando no sea este el propósito y la ubicación de las acciones evaluativas cabe tener presente que para los actores participantes en alguna de las instancias del proceso, éste adquiere esta función simbólica.
- **función política:** tal como se señalara en el apartado anterior una de las funciones más importantes de la evaluación es su carácter instrumental central como soporte para los procesos de toma de decisiones. Esta función es claramente política ya que la evaluación adquiere un rol sustantivo como retroalimentación de los procesos de planificación y la toma de decisiones sobre la ejecución y el desempeño de los programas y proyectos.
- **función de conocimiento:** en la definición misma de evaluación y en la descripción de sus componentes se identifica como central el rol de la evaluación en tanto herramienta que permite ampliar la comprensión de los procesos complejos; en este sentido la búsqueda de indicios en forma sistemática implica necesariamente el incremento en el conocimiento y la comprensión de los objetos de evaluación.
- **función de mejoramiento:** en forma complementaria con la función de conocimiento y la identificada como función política, esta función destaca el aspecto instrumental de la evaluación en tanto permite orientar la toma de decisiones hacia la mejora de los procesos o fenómenos objeto de evaluación. En la medida que se posibilita una mayor comprensión de los componentes presentes es factible dirigir las acciones hacia el mejoramiento en términos de efectividad, eficiencia, eficacia, pertinencia y/o viabilidad de las acciones propuestas.

• **función de desarrollo de capacidades:** con carácter secundario, ya que no forma parte de los objetivos centrales de cualquier acción evaluativa, los procesos de evaluación a través de sus exigencias técnicas y metodológicas desempeñan una importante función en términos de promover el desarrollo de competencias muy valiosas. Si se aprovechan adecuadamente las instancias de evaluación, éstas contribuyen a incrementar el desarrollo de dispositivos técnicos institucionales valiosos y poco estimulados habitualmente. Estas competencias se refieren por ejemplo a la práctica sistemática de observaciones y mediciones, de registro de información, de desarrollo de marcos analíticos e interpretativos de la información, de inclusión de la información en los procesos de gestión, de desarrollo de instrumentos para la recolección de información, etc.

b) Para qué se evalúa

La pregunta por la finalidad de la evaluación constituye uno de los ejes centrales junto con la definición del objeto a ser evaluado. En términos generales y en función de las definiciones adoptadas permiten identificar algunas finalidades principales de todo proceso evaluativo: diagnóstico, pronóstico, selección y acreditación estas finalidades no necesariamente se plantean como excluyentes pero sí determinan opciones metodológicas diferenciales.

La finalidad de diagnóstico enfatiza los componentes vinculados con la producción sistemática de información calificada con el objeto de orientar la toma de decisiones, la gestión.

La finalidad de pronóstico enfatiza el valor predictivo que pueda tener la información que se produce, es decir las acciones evaluativas persiguen como propósito la producción de información con alto potencial anticipatorio y explicativo sobre los fenómenos o procesos objetos de evaluación.

La finalidad de la selección pone el énfasis en la utilización que tiene la información producida por la evaluación con propósitos de selección, un ejemplo claro de esto son los exámenes de ingreso a diferentes instituciones educativas cuyos aspirantes superan el número de vacantes disponibles. En estas circunstancias se opta por alguna estrategia de evaluación que con frecuencia se justifica adjudicándole un valor pronóstico a estos resultados.

La finalidad de acreditación es la que más se vincula con este valor social – simbólico que tiene la evaluación. En estos casos en énfasis está puesto en las consecuencias que los resultados de la evaluación tienen para el individuo o la institución objeto de evaluación ya que de su resultado depende la continuidad de los estudios para un sujeto o la interrupción parcial de su carrera escolar, etc..

c) Los ámbitos de la evaluación

Si bien las imágenes más frecuentes sobre la evaluación se refieren a la misma aplicada a los alumnos, es importante tener en cuenta que los puntos centrales señalados en este trabajo son también aplicables a diferentes ámbitos de la evaluación educativa.

Frente a la tarea, es necesario precisar –junto con la finalidad de la evaluación – el ámbito dónde la misma tendrá lugar y procurar definir las herramientas metodológicas consideradas más adecuadas para ese ámbito, tanto en las tareas de recolección cuanto de sistematización y análisis de la información.

De este modo es posible diferenciar la evaluación:

- de los aprendizajes.
- de las instituciones.
- del sistema educativo.
- de los programas o proyectos.

d) A modo de síntesis

Todo proceso de evaluación exitoso debe observar algunos requerimientos indispensables o atender a cierto conjunto de premisas.

En primera instancia debe entenderse que toda acción evaluativa es una forma particular de intervención en la realidad. Cualquiera que sea la/s estrategia/s metodológica/s adoptada/s los objetos de evaluación sufren algún tipo de modificación como consecuencia de esta intervención. Este es un elemento que debe tener presente el evaluador, por una parte para no producir interferencias en demasía en el desarrollo habitual de las acciones y por la otra

Para ponderar este efecto en la elección de las estrategias metodológicas – por ejemplo – y el balance general de los resultados de la evaluación.

En segunda instancia toda evaluación, por su naturaleza, requiere de criterios establecidos respecto de los cuales se formulen los juicios valorativos. Ya se señaló anteriormente que el componente valorativo es uno de los elementos diferenciales de la evaluación respecto de cualquier otro tipo de indagación y en este sentido la formulación, definición o construcción de criterios resulta un requerimiento ineludible.

En tercer lugar y teniendo en cuenta lo señalado en apartados anteriores, dado que toda acción evaluativa se sustenta en la producción de información y en la búsqueda de indicios sobre aquellos procesos o fenómenos

no visibles en forma simple, es imposible plantear una evaluación que abarque todos los aspectos a considerar, que incluya todos los componentes de un programa o proyectos, que indague sobre todos los procesos de gestión, etc. Por el contrario cualquier planteo al respecto implica siempre ciertos límites conceptuales, metodológicos, operativos, pero en todos los casos son límites que suponen alguna forma de acotamiento de los alcances de la evaluación.

La información constituye el insumo básico sobre el que se producen los procesos de evaluación por lo tanto el rigor técnico dirigido hacia la selección las fuentes, las formas de recolección, registro, procesamiento y análisis, así como la confiabilidad y validez resultan exigencias insoslayables a la hora de llevar a cabo cualquier tipo de evaluación.

Las características mismas de los procesos evaluativos traen como consecuencia una notable movilización por parte de todos los sujetos de que una u otra manera se ven involucrados en ellos. Por ello es necesario considerar en términos generales quienes son los sujetos alcanzados de una y otra forma por las acciones evaluativas y generar entre ellos cierta base de consenso y aceptación. Esto contribuye a evitar la resistencia propia que generan este tipo de acciones en parte debido a la endeble cultura evaluativa que en general se observa. La comunicación amplia sobre las finalidades de las acciones de evaluación, los procedimientos predominantes, la utilización de los resultados, la difusión de los mismos son estrategias que promueven una mayor aceptación entre los involucrados.

La relevancia y la oportunidad son, por último, dos requisitos básicos para que las acciones de evaluación tengan significación. La relevancia se refiere a la centralidad de el/los objetos de evaluación, es decir, la complejidad de un proceso de evaluación se justifica en la medida en que se lleve a cabo alrededor de problemas, aspectos o fenómenos relevantes. En el mismo sentido la oportunidad es la cualidad de la ubicación temporal adecuada del proceso de evaluación y en especial de sus resultados. Un proceso de evaluación exitoso lo es en buena medida en tanto sus resultados se encuentren disponibles en el momento adecuado, para incidir en la toma de decisiones, en los procesos de gestión en forma oportuna.

LA EVALUACIÓN EDUCATIVA, UN CRITERIO CONSOLIDADO.

Introducción

A tono con la tendencia internacional, en la Argentina se va consolidando el criterio de establecer sistemas de medición de la calidad de la enseñanza, para evaluar el nivel de educación que se imparte en las aulas.

En este informe se plantea la evolución que han tenido en el sistema educativo los instrumentos que apuntan a medir la calidad de la enseñanza. Más que un análisis exhaustivo de los resultados, se ofrecen aportes para intentar explicar las causas de su surgimiento y expansión, las metodologías empleadas y las distintas concepciones acerca de para qué y por qué evaluar.

Los sistemas de evaluación en América Latina y en Argentina

Prácticamente todos los países de América latina han iniciado, con distinto grado de aplicación, algún tipo de programa de evaluación del sistema educativo. Así lo señala un informe del PREAL (Programa de Promoción de la Reforma Educativa en América latina y el Caribe), que ubica a México y Chile entre los países de la región que más experiencia han desarrollado en el tema.

El trabajo, realizado por el especialista norteamericano Lawrence Wolff, asesor en educación del Banco Interamericano de Desarrollo (BID) y publicado por PREAL en julio de 1998, sostiene que la Argentina comenzó tardíamente –en 1993– con la aplicación de pruebas de evaluación de la calidad. Considera que el programa argentino está bien concebido y tiene una estrecha relación con la estrategia de descentralización de los servicios educativos y con el mejoramiento de la calidad de la enseñanza. Agrega que si bien se perciben logros, aún no se han establecido “objetivos de aprendizaje claros a nivel nacional”.

Sólo ahora, al cabo de varios años, se iniciaron “esfuerzos sistemáticos por compatibilizar el currículum, los textos escolares y la pedagogía utilizada en la sala de clases”.

Los operativos de evaluación educativa en la Argentina comenzaron en 1993, con los exámenes de lengua y matemática que rindieron alumnos de séptimo grado y quinto año del secundario. Las 38.000 pruebas administradas tuvieron carácter muestral (no incluyó a todos los alumnos, sino a una parte representativa de esos niveles de enseñanza) y reflejaron un promedio de 6,14 en lengua y de 4,63 en matemática entre los que cursaban el último año de la enseñanza media.

Se extendieron cada año a más alumnos y a partir de 1997 el examen de finalización del secundario se amplió a todos los estudiantes de ese nivel. Así, unos 280.000

jóvenes rinden anualmente las pruebas de lengua y matemática, cuyos resultados no tienen incidencia en el promedio final del alumno, aunque sirven para conocer el nivel de educación que se imparte en las aulas. Las pruebas nacieron con la intención de promover un certificado nacional de estudios básicos y la pretensión de que sus resultados influyeran en el acceso a la enseñanza superior. Pero tales ideas han sido por ahora dejadas de lado y se delegó en cada jurisdicción provincial la decisión de que las notas de los alumnos figuren en el boletín de calificaciones.

Mientras se conservan en forma muestral las evaluaciones en distintas etapas del nivel primario, los mayores esfuerzos se han concentrado en el desarrollo de las pruebas de finalización del secundario.

También los organismos internacionales, agrega Tiana, se sumaron a esta corriente y pusieron en marcha programas vinculados con el desarrollo de las políticas de evaluación educativa. Ejemplo de ello son la Unesco, la Unión Europea, la Organización para la Cooperación y el Desarrollo Económico (OCDE) y la Organización de Estados Iberoamericanos (OEI).

Creadas inicialmente para medir el desempeño de los alumnos en lengua y matemática, las pruebas de evaluación se extendieron en los últimos años a otras áreas (física, química, historia, geografía, educación cívica) y a partir del año 2000 se incorporaron tests internacionales, con pruebas reguladas por organizaciones de alcance mundial.

En consideración de los especialistas, Chile cuenta con uno de los sistemas de evaluación más amplios y mejor administrados de América latina. El informe de Lawrence Wolf destaca “haber demostrado el mayor compromiso de largo plazo con el desarrollo de evaluaciones”. El programa se inició en 1980 y desde 1988 se denomina Sistema de Medición de la Calidad de la Educación. La herramienta sirvió para acompañar el plan de reforma que promueve la descentralización.

De acuerdo con ese trabajo, México es el país latinoamericano con más larga experiencia en las evaluaciones. Pero se caracteriza por la renuencia de sus autoridades a divulgar sus resultados. A la medición de los conocimientos y habilidades de 2,8 millones de estudiantes, acompañan una evaluación sistemática de la capacidad de los profesores.

Brasil comenzó a desarrollar en 1990 un sistema de evaluación nacional, pero sólo a partir de 1995 comenzó a tomar muestras. Desde esa fecha se inició la evaluación de la enseñanza superior, el primero en su tipo en la región. El gobierno brasileño estimula a los Estados y municipalidades a iniciar sus propias evaluaciones.

El informe de Wolff afirma que en América latina las asociaciones gremiales de profesores tienden a ser observadores pasivos, e incluso, se oponen a las evaluaciones. El especialista del BID concluye que si se desea que éstas tengan impacto en la educación es preciso incluir desde el comienzo a los profesores en ese proceso. Propone incluir otras herramientas de medición, como los estudios de deserción y repitencia, mediciones de los insumos escolares y estimaciones de los recursos mínimos que cada escuela debería tener, observaciones sistemáticas de los procesos escolares y estudios del desempeño de los egresados en el mercado laboral.

Una de las herramientas que contribuyó a consolidar en todo el mundo el sistema de evaluación de la enseñanza es el Third International Mathematics and Science Study (TIMSS), que compara y explica el aprendizaje en ciencia y matemática en 41 países. La Argentina se sumó a estos exámenes, rendidos por alumnos de primer año del secundario, en el año 2000.

La participación de la Argentina en los tests internacionales comenzó con el TIMSS y se extendió a otros programas. El objetivo central es comparar la educación nacional con la de otros países, incorporar e intercambiar experiencias y capacitar a los equipos técnicos.

Además del mencionado TIMSS, las pruebas internacionales incorporadas por la Argentina corresponden a tests diseñados por la IEA (Asociación Internacional para la Evaluación del Logro Educativo) y la OCDE (Organización para la Cooperación y el Desarrollo Económico). Son las siguientes:

- **PIRLS** (Estudio sobre los avances internacionales en lectura y alfabetización). Fue diseñado para proveer información sobre la habilidad de lectura y comprensión de los alumnos de cuarto grado (9 a 10 años).
- **CIVICS** (Estudio internacional de educación cívica). Está destinado a alumnos de 14 y 15 años, para investigar la forma en que los jóvenes son preparados para incorporar los valores cívicos y la cultura democrática y aprendan a tomar parte en los asuntos públicos.
- **PISA** (Programa para la evaluación internacional de los estudiantes). Es una encuesta sobre habilidades y conocimientos dirigidos a chicos de 15 años. Consiste en ejercicios sobre las áreas de lectura, matemáticas y ciencias y los alumnos deben responder un cuestionario sobre su desarrollo personal y actitudes.

a) Laboratorio de la UNESCO

En octubre de 2000, la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC) difundió los resultados de un estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos de tercer y cuarto grado de la educación básica. La investigación fue realizada por

el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación y comprendió a 54.000 alumnos de 13 países.

El estudio analiza el rendimiento de los alumnos en la región y revela que el promedio de logro de los estudiantes es más bajo de lo esperado.

La mayoría de los estudiantes realiza una comprensión fragmentaria de los textos que leen, reconocen las palabras incluidas en un texto pero no consiguen determinar por qué se dice lo que se dice o para qué se dice. Ello podría indicar que a los niños se les enseña a decodificar, a traducir las palabras escritas al lenguaje oral, pero sin entender el significado del texto, ni interpretar lo que leen.

Los resultados en matemática son generalizadamente más bajos y desiguales. Los alumnos no asimilan los conocimientos ni desarrollan las competencias en la asignatura.

Como dato ilustrativo se añade que los estudiantes cubanos alcanzaron los mayores puntajes en lenguaje y matemática y que las niñas alcanzan mejores logros en lenguaje y ligeramente menores en matemática. (Nota de la redacción: Argentina obtuvo el 2º lugar en ambas asignaturas, lo que la ubicó a la cabeza de los demás países participantes).

La elaboración de estos programas es muestra del creciente interés suscitado en el nivel internacional por la evaluación de los sistemas educativos. Según el especialista Tiana, ello ha producido como efecto una "rápida evolución de la evaluación entendida como disciplina científica y como práctica profesional".

Entre otras causas que explican ese interés, Tiana identifica la creciente demanda social de información sobre la educación que se aprecia en los distintos países. Una suerte de rendición de cuentas. "En épocas de crisis, existe la necesidad de establecer prioridades en la asignación de recursos y para ello resulta inevitable fijar criterios de comparación lo más objetivos posible", es el pensamiento del profesor español.

Para Tiana, un requisito ineludible para cualquier política de evaluación es su credibilidad. Ello se vincula directamente con la independencia institucional de los mecanismos de evaluación.

Otros criterios esenciales son la participación de los sectores implicados y el desarrollo gradual de las políticas de evaluación, adaptadas a las circunstancias y a los contextos determinados. "Es absurdo pretender desarrollar políticas de evaluación a partir de la importación de modelos foráneos", es uno de los consejos que el experto español transmite a sus colaboradores y discípulos.

b) Para qué evaluar

Muchos autores coinciden en que uno de los objetivos primordiales de la evaluación educativa es informar y orientar la toma de decisiones. Esa idea, que habla de una "utilización instrumental de los resultados", es el concepto que predominó en las primeras etapas del desarrollo histórico de la evaluación. El autor W. J. Pophan

recuerda que en los años 70 existía la creencia dominante de que las evaluaciones educativas debían constituir el elemento singular más importante para la toma de decisiones.

Hoy, sin embargo, se tiende a considerar que la relación existente entre evaluación y toma de decisiones es más compleja. "La evaluación es uno de los elementos que contribuyen a la toma de decisiones. Pero no el único", advierte Tiana en sus seminarios y conferencias.

Y explica que la concepción puramente instrumental de la evaluación se fue debilitando con el tiempo y abrió paso a la concepción conceptual o iluminativa, según la cual la evaluación no tendría como función principal la de ofrecer elementos para la toma de decisiones; también contribuiría a mejorar el conocimiento de los procesos educativos. "En vez de pensar en un impacto inmediato y directo, la concepción iluminativa considera los efectos diferidos e indirectos de la evaluación, en plazos de tiempo más largos que los habituales", explica el especialista español.

Algunos autores añaden una tercera concepción. Rossi y Freeman (1993) hablan de una utilización persuasiva y sostienen que la evaluación desempeñaría una función eminentemente argumentativa al servicio del discurso político, sea éste de gobierno o de oposición. Muchas veces se corre el riesgo de que los gobiernos de turno "hagan públicas sólo aquellas partes del informe que tienen un carácter halagador", advierte el especialista Miguel Ángel Santos, citado por Tiana en un documento sobre "Tratamiento y usos de la información en evaluación".

c) Hacia dónde vamos

En el documento "Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América latina?", elaborado en marzo de 2000 por un equipo que dirigió Pedro Ravela y publicado también por PREAL, se señala que muchos países ingresaron en una fase de revisión de lo hecho hasta el momento y comenzaron a considerar nuevas alternativas.

Ello responde, se explica en el informe, a los siguientes motivos:

Poco aprovechamiento de la información producida por los sistemas de evaluación. Se percibe un insuficiente impacto en el sistema educativo.

Insuficiente calidad y capacidad de evaluación de aprendizajes complejos en las pruebas que están siendo aplicadas.

Debilidades técnicas en los procesos de desarrollo y validación de los instrumentos de medición.

En el citado documento se reconoce que la instalación de los sistemas nacionales de evaluación ha dado pasos importantes, pero se considera necesario definir con qué estrategias se espera aplicarlos para que tengan algún impacto en la mejora de los aprendizajes.

Para ello, los evaluadores deben tener claro si prefieren una evaluación con consecuencias directas para las escuelas y maestros –que determinen la aprobación o reprobación de los alumnos- o si esperan cumplir una

función fundamentalmente informativa, si desean contar con una información exhaustiva sobre las competencias y conocimientos de los alumnos o si prefieren producir información menos detallada, limitada a cada establecimiento.

Ante la necesidad de capacitar cuadros técnicos para una experiencia novedosa como la evaluación del sistema educativo y ante la escasa "masa crítica" existente en la región, en el informe se sugiere facilitar el contacto con especialistas de la comunidad internacional.

También se considera conveniente considerar las estrategias de uso y difusión de los resultados de las evaluaciones.

Sitiografía

- www.iipe.unesco.org: Oficina Internacional de Educación de la UNESCO.
- www.iipe-buenosaires.org.ar: Sitio del Instituto Internacional de Planeamiento de la Educación (IIEPE), centro de formación e investigación creado por la UNESCO.
- www.unesco.cl/home.htm: Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC)
- www.me.gov.ar: Sitio del Ministerio de Educación de la Nación.
- www.oei.es: Organización de Estados Iberoamericanos (OEI).
- www.preal.cl: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PERAL).
- www.iea.nl/Home/home.html: International Association for Evaluation of the Educational Achievement (IEA) y acceso a los sitios de los programas TIMSS, PIRLS y Civics.
- www.eclac.org: Comisión Económica para América Latina y el Caribe (CEPAL).

Bibliografía

- "¿Cómo avanzar en la evaluación de aprendizajes en América Latina?", por Pedro Ravela (editor), Richard Wolfe, Gilbert Valverde y Juan Manuel Esquivel. Grupo de Trabajo sobre Estándares y Evaluación GRADE/PREAL. Marzo de 2000.
- "La educación secundaria. ¿Cambio o inmutabilidad?". Análisis y debate de procesos europeos y latinoamericanos contemporáneos. Org. Cecilia Braslavsky. Santillana. Abril 2001.

- "Tratamiento y usos de la información en evaluación", por Alejandro Tiana Ferrer. Universidad Nacional de Educación a Distancia (UNED), de España.
- "Las evaluaciones educacionales en América Latina: avance actual y futuros desafíos", por Lawrence Wolf. PREAL. Julio de 1998.
- III Operativo Nacional de Finalización del Secundario 1999. Ministerio de Educación de la Nación.
- "La evaluación de los sistemas educativos", por Alejandro Tiana Ferrer. Revista Iberoamericana de Educación de la OEI. Número 10. Enero/Abril 1996.
- De Ketele, J.M. y Roegiers, X. (1995) Funciones y campos de aplicación de la recogida de
- Información. La Muralla, Madrid.
- Arregui, Patricia. Sistemas de determinación y evaluación de metas de logros de aprendizaje escolar como instrumentos para mejorar la calidad, la equidad y la responsabilización en los procesos educativos en América Latina.
- Seminario sobre Prospectivas de la Educación en América Latina y el Caribe,
- Chile, 23 al 25 de agosto de 2000.UNESCO.
- Ravela, Pedro. (Agosto, 2001)¿Cómo presentan sus Resultados los Sistemas nacionales de Evaluación Educativa en América Latina? PREAL.

EVALUACIÓN NACIONAL Y EVALUACIONES INTERNACIONALES

A continuación se presenta un cuadro comparativo entre el Estudio de Evaluación Nacional y los dos Estudios Internacionales que se aplican en la actualidad en nuestro país.

	ONE	SERCE	PISA
OBJETIVOS Y CARACTERÍSTICAS	<p>Es el Operativo Nacional de Evaluación 2007 sobre Lengua, Matemática, Ciencias Sociales y Ciencias Naturales en alumnos de 3°, 6° de la Primaria y 2°/3°, 5°/6° de la Secundaria, y Cuestionarios de Contexto destinados a los Directores de Primaria y Secundaria, a los Alumnos de 6°, 2°/3° y 5°/6° y a los Docentes de 6° de Ciencias Naturales. Cabe destacar que en el año 2007 no se aplicaron las pruebas de Ciencias Naturales y Ciencias Sociales a los alumnos de la Secundaria.</p> <ul style="list-style-type: none"> ■ De acuerdo con la Nueva Ley de Educación Nacional, la aplicación del ONE tiene entre sus propósitos: <ul style="list-style-type: none"> -Evaluar determinados desempeños alcanzados por los alumnos en el país, a través de sus capacidades y de sus contenidos, en los diferentes niveles y áreas. - Aportar los insumos estadísticos y pedagógicos, a partir de la "evaluación de los aprendizajes de los alumnos y de los factores asociados", para mejorar el desarrollo de políticas educativas sustentadas en la realidad de las aulas. ■ Es una propuesta de evaluación con base en los Contenidos Básicos Comunes (CBC), los diseños curriculares jurisdiccionales, los Núcleos de Aprendizajes Prioritarios (NAP) y los resultados de los Operativos Nacionales de Evaluación. 	<p>El Segundo Estudio Regional Comparativo y Explicativo (SERCE 2006) sobre Lenguaje, Matemática, Ciencias y Factores Asociados en alumnos de 3° y 6° año de E.G.B de 16 países de la región.</p> <ul style="list-style-type: none"> ■ El SERCE es un producto de una construcción colectiva y participativa de los países latinoamericanos que conforman la red del Laboratorio Latinoamericano de Evaluación de Calidad Educativa (LLECE). ■ El estudio busca aportar información a los sistemas educativos de la región. ■ Desarrolla una propuesta de evaluación propia y con base en los diseños curriculares de los países participantes de la Región. ■ El enfoque desde el que los conocimientos y capacidades son evaluados se corresponden con las habilidades para la vida en la evaluación de Lenguaje, Matemática y Ciencias, y el análisis curricular SERCE. 	<p>El Estudio PISA - Programme for International Student Assessment (Programa para la Evaluación Internacional de Estudiantes) es una evaluación estandarizada diseñada y coordinada por la OCDE (Organización para la Cooperación y el Desarrollo Económico).</p> <p>Esta evaluación se diseña a partir de un conjunto de estándares de contenido y desempeño, elaborados por un grupo de especialistas en diferentes áreas de conocimiento (Comprensión Lectora, Matemática y Ciencias), que se consideran imprescindibles para que los jóvenes puedan desenvolverse con éxito en la sociedad y en el mundo laboral de hoy.</p> <p>Evalúa también la capacidad de los estudiantes de organizar y regular su propio aprendizaje (aprendizaje autorregulado), la capacidad de resolver situaciones problemáticas en contextos interdisciplinarios y de la vida cotidiana (capacidad de resolver problemas).</p> <p>Para ello, los alumnos deben ser capaces de reconocer un problema, definir su naturaleza, usar el conocimiento para delinear una estrategia de resolución, ajustar la solución que mejor se adapte al problema y comunicar a otros la solución hallada.</p>

	ONE	SERCE	PISA
POBLACIÓN EVALUADA	<p>Todas las escuelas que participan de la muestra en el país han sido elegidas por la DINIECE de manera aleatoria, es decir, al azar mediante procedimientos estadísticos variados para este tipo de estudio.</p> <ul style="list-style-type: none"> ■ En las escuelas seleccionadas, se evaluará a los grupos de alumnos de 3° y 6° años de Primaria y de 2°/3° y 5°/6° años de Secundaria. Se realizará una “extensión al ámbito rural” evaluando a grupos de alumnos de 3° y 6° años de Primaria, en aproximadamente 2000 escuelas rurales. <ul style="list-style-type: none"> - Escuelas Primarias: 4000, aprox. - Escuelas Secundarias: 2000, aprox. - Alumnos de 3°: 80.000, aprox. - Alumnos de 6°: 75.000, aprox. - Alumnos de 2°/3°: 48.000, aprox. - Alumnos de 5°/6°: 48.000, aprox. 	<p>Todas las escuelas que participaron de la muestra en nuestro país han sido elegidas por la UNESCO de manera aleatoria, es decir, al azar mediante procedimientos estadísticos variados para este tipo de estudio. En las escuelas seleccionadas, se evaluaron a los grupos de alumnos de 3° y 6° años de la escuela Primaria.</p> <ul style="list-style-type: none"> ■ Escuelas del país: 169 ■ Alumnos de 3° año: 7.202 ■ Alumnos de 6° año: 7.014 ■ Secciones de 3° año: 309 ■ Secciones de 6° año: 307 	<p>La población elegida para evaluar es la de los estudiantes que tienen 15 años, ya que ésta es la edad en la que la mayoría de ellos finaliza su escolaridad obligatoria en los países miembros.</p> <p>Esta evaluación es administrada entre 4500 y 10000 estudiantes en cada país.</p> <p>La delimitación de la población a estudiar en términos de edad, y no de curso, proporciona un punto estándar en todos los países, aunque tiene el inconveniente de no tener un referente educativo específico en un determinado grado de escolaridad.</p>
CICLO DE EVALUACIÓN	<p>La Dirección Nacional de Información y Evaluación de la Calidad Educativa -DINIECE- aplica los Operativos Nacionales de Evaluación -ONE- desde el año 1993 hasta el año 2000 en forma ininterrumpida. Luego de la crisis económica y política los operativos de evaluación se realizaron en el 2002 y 2003. A partir del 2003 se decide realizar operativos de calidad cada dos años, en el año 2005 y 2007 se realizaron los últimos operativos de evaluación.</p>	<p>El Primer Estudio Internacional Comparativo (1997) sobre Lenguaje, Matemática, y Factores Asociados en alumnos de 3° y 4° Año de E.G.B de 13 de países de la región.</p> <p>El Segundo Estudio Regional Comparativo y Explicativo (2006) sobre Lenguaje, Matemática, Ciencias y Factores Asociados en alumnos de 3° y 6° año de E.G.B de 16 países de la región.</p>	<p>La primera evaluación tuvo lugar en el año 2000 con los primeros resultados publicados en el 2001 y a partir de entonces se continúa en ciclos de tres años. Esta aplicación es de carácter cíclico para poder seguir el rendimiento de los estudiantes de 15 años y establecer relaciones longitudinales con respecto al desempeño académico de esas poblaciones.</p>
PAÍSES Y JURISDICCIONES	<p>Todas las jurisdicciones del país</p>	<p>SERCE 2006/2007: 17 países (Argentina, Brasil, Colombia, Costa Rica, Cuba Chile República Dominicana, Ecuador, El Salvador, Guatemala, Nicaragua, México, Panamá, Paraguay, Perú, Uruguay y el Estado mexicano de Nuevo León)</p>	<p>PISA 2000: 43 países PISA 2003: 41 países. PISA 2006: 57 países. PISA 2009: 62 países.</p>
LOS INSTRUMENTOS	<p>(*) Ver Anexo</p>	<p>(*) Ver Anexo</p>	<p>(*) Ver Anexo</p>

ANEXO 1

ONE

Estructura y características de los instrumentos

Tipo	Nombre	Cuadernillos	Forma de administración	¿Quién lo responde?
Pruebas de conocimiento	Lengua	Del 1 al 3	Sesión colectiva	Alumnos de 3°, 6°, 2°/3° y 5°/6°
	Matemática	Del 1 al 3		Alumnos de 3°, 6°, 2°/3° y 5°/6°
	Ciencias Sociales	Del 1 al 3 ó 4		Alumnos de 3°, 6°
	Ciencias Naturales	Del 1 al 3 ó 4		Alumnos de 3°, 6°
Cuestionarios de Contexto	Del Alumno de 6°	Único	Autoadministrada	Alumnos de 6°
	Del Alumno de 9°	Único	Autoadministrada	Alumnos de 2°/3°
	Del Alumno de 12°	Único	Autoadministrada	Alumnos de 5°/6°
	Del Director de Primaria	Único	Autoadministrada	Director o miembro del equipo directivo
	Del Director de Secundaria	Único	Autoadministrada	Director o miembro del equipo directivo
	Del Docente de 6° de Ciencias Naturales	Único	Autoadministrada	Todos los Docentes de los alumnos evaluados de 6° de Ciencias Naturales

SERCE

La siguiente Tabla resume la lista de instrumentos que se aplicaron en cada escuela:

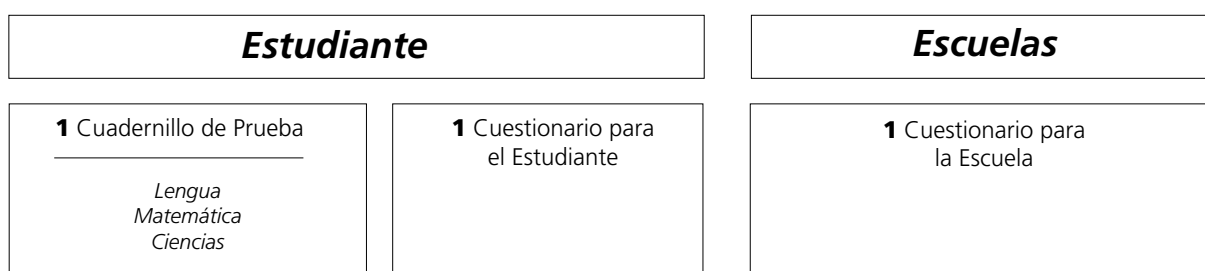
Tipo	Nombre	Cuadernillos	Código	Forma de Administración	¿Quién lo responde?
Pruebas de Logro	Lectura	Del 1 al 6	DL3 – DL6	Sesión colectiva	Alumnos de 3 y 6
	Escritura	Del 1 al 2	DE3 – DE6		Alumnos de 3 y 6
	Matemática	Del 1 al 6	DM3 – DM6		Alumnos de 3 y 6
	Ciencias	Del 1 al 6	DC6		Alumnos de 6
Cuestionarios de Factores Asociados	Del alumno de 3	Único	QA3		Alumnos de 3
	Del alumno de 6	Único	QA6		Alumnos de 6
	Del docente	Único	QP	Autoadministrada	Docentes de lenguaje y matemática del aula evaluada
	De enseñanza	Cinco	QL3, QL6, QM3, QM6, QC6	Autoadministrada	Docente del área respectiva del aula evaluada
	De la familia	Único	QF	Autoadministrada / Entrevista	Padres o tutores de los alumnos de 3 y 6 evaluados
	Del directivo	Único	QD	Autoadministrada	Director o miembro del equipo directivo
	Ficha de empadronamiento	Único	FE	Entrevista	Aplicador con el apoyo del director

PISA

Estructura y características de los instrumentos

Los instrumentos son de dos tipos:

- Pruebas de rendimiento
- Cuestionarios de contexto



Pruebas de rendimiento

Las pruebas de rendimiento se construyen a partir de los textos, ejercicios y preguntas aportadas por los países miembros de la OCDE, las que son revisadas y seleccionadas por los órganos de la Agencia Evaluadora, responsables del proyecto.

Las pruebas están constituidas por ítem de tres tipos:

- Cerrados con cuatro o cinco opciones de respuesta para elegir la correcta.
- Abiertos de respuesta corta y ejercicios abiertos de respuesta extendida, en los que el alumno tiene que construir y desarrollar la respuesta.
- Actitudinales, a través de ellos se obtiene importante información sobre contenidos actitudinales, considerados centrales en la educación científica. Los resultados de estos ítems no son contemplados en el puntaje final del rendimiento académico.

Cuestionarios de contexto

En Argentina se aplican dos cuestionarios, uno dirigido al estudiante de 15 años y otro al Establecimiento Educativo, completado por el director del mismo.

PISA propone otros instrumentos de recolección de datos, como ser: Cuestionario para padres y Cuestionario de Educación Tecnológica, que hasta el momento Argentina no ha aplicado.

A través de la recolección de datos adicionales sobre los alumnos (actitudes y conductas) y el entorno educativo (condiciones de enseñanza y de aprendizaje) PISA 2006 puede identificar factores sociales, culturales, económicos y educativos que se cruzan con el desempeño de los alumnos.

Cuestionario para el alumno

En cada ciclo el estudio PISA recolecta información demográfica básica como un componente núcleo que replica las preguntas claves de los ciclos anteriores: género, lengua materna, nivel socioeconómico de los alumnos (tomando como elementos importantes el nivel de educación de los padres y la situación económica familiar), el desarrollo educativo de los alumnos, la disponibilidad y el uso de los recursos educativos en el hogar y en la escuela, y las prácticas de enseñanza, incluyendo aspectos tales como los relativos al currículum, el tiempo dedicado a la escuela y a las tareas escolares, etc.

Cuestionario para el establecimiento educativo

El componente común recopila información sobre la situación geográfica de la escuela, su contexto socioeconómico, incluyendo los recursos educativos del establecimiento y su entorno, el tamaño de la escuela y de sus clases, la estructura de su personal, las prácticas escolares en el nivel de la escuela y la organización escolar.

Por otra parte, este cuestionario releva información sobre aspectos de la rendición de cuentas del Establecimiento ante los padres, sobre la participación y promoción de actividades científicas en las que participen los alumnos y docentes, si se abordan temas relacionados con el medio ambiente, sobre la orientación y perspectivas que tienen los jóvenes sobre su futuro y sobre la promoción de las carreras relacionadas con la ciencia por parte de los actores escolares.

“¿QUÉ PRETENDEMOS EVALUAR, QUÉ EVALUAMOS Y QUÉ CONCLUSIONES PODEMOS EXTRAER DE LA EVALUACIÓN?”

Los años '90 se caracterizaron por la puesta en marcha y el desarrollo de un buen número de reformas educativas en los países latinoamericanos. Una vez superada la desconfianza hacia la educación que caracterizó a los años '70 y '80, los gobiernos de la región se lanzaron a una intensa actividad reformadora, que se dejó sentir en muchos países y que abarcó diversas esferas de la realidad educativa (Gajardo, 1999).

Las reformas emprendidas en esos años persiguieron varios objetivos, entre los que cabe destacar la extensión de la escolarización (especialmente en la enseñanza secundaria), la mejora de la calidad de la educación impartida y el aumento de la equidad en la distribución de los servicios educativos. Aunque no sea éste el lugar para hacer un balance de la década, no cabe duda de que los objetivos previstos se alcanzaron de manera muy desigual, consiguiéndose algunos logros destacables en el primer sentido de los mencionados, pero logrando resultados mucho más modestos en los otros dos. La situación es tal que un reciente informe internacional sobre el estado de la educación en los países de América Latina adoptaba como título el lema *Quedándonos atrás*, expresando de ese modo las carencias que aún aquejan a los sistemas educativos de la región (Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe, 2001).

El desarrollo de políticas nacionales de evaluación

Las reformas de los años '90 afectaron a diversos aspectos de la realidad educativa, que conviene detallar. Entre las orientaciones adoptadas por dichos procesos de reforma, y aun sin ánimo de exhaustividad, pueden distinguirse varias líneas de actuación que, si bien no se dieron siempre de manera simultánea, estuvieron presentes de uno u otro modo en las diversas iniciativas nacionales:

- Un primer aspecto en que se avanzó considerablemente, aunque no sin problemas, fue en la redistribución de las competencias y responsabilidades en materia de educación, lo que supuso la puesta en marcha de políticas de descentralización y de autonomía escolar, acompañadas en ocasiones de procesos de privatización, cuyos

efectos fueron ambivalentes (Winkler y Gershberg, 2000).

- Una segunda orientación de las reformas, especialmente influyente en muchos países de la región, consistió en la introducción de cambios en la organización curricular, que llevaron generalmente aparejada una revisión de los planteamientos fundamentales en que se inspira la definición y la construcción del currículo.
- Un tercer dominio de acción de las reformas tuvo que ver con los procesos de formación inicial y capacitación del profesorado, considerado habitualmente pieza clave de las reformas, pero no siempre tan atendido como debiera en su papel de agente transformador.
- Un cuarto ámbito de reforma estuvo concretamente relacionado con la educación secundaria o media, que se ha convertido en la pieza clave y más conflictiva de los sistemas educativos y cuya expansión se ha ido también produciendo en América Latina en las últimas décadas (Braslavsky, 2001).
- Por fin, una última orientación tuvo que ver con el desarrollo de mecanismos y modelos de evaluación; dicho de otro modo, con la consideración de la evaluación, como un poderoso instrumento al servicio de la gestión de la educación y de la mejora de su calidad (Tiana, 1996).

Esta última orientación de las reformas de los '90 cobra especial relevancia desde la perspectiva que se adopta en este trabajo, pues es precisamente el impacto de esos mecanismos de evaluación lo que está sometido a crítica y discusión. Por este motivo, merece la pena realizar un par de comentarios adicionales sobre sus rasgos más significativos.

Una primera observación que cabe destacar desde este punto de vista es que la mayoría de los países de la región pusieron en marcha sus propios sistemas nacionales de evaluación a lo largo de los años '90, generalmente en el marco de procesos más amplios de reforma educativa. Con configuraciones institucionales muy diferentes entre sí y con ámbitos de actuación diversos, fueron varios los gobiernos que se sumaron a la experiencia pionera que otros países como Chile venían desarrollando ya con anterioridad. Argentina, México, Colombia, Bolivia o Brasil, por no citar sino algunos casos destacados, dieron pasos decididos en esa dirección,

de manera que al final de la década eran mayoría los países latinoamericanos que contaban con algún tipo de sistema nacional de evaluación (Ravela, 2001).

Sin embargo, el indudable interés por la evaluación de la educación que se despertó en muchos países no marchó asociado sino mucho más tardíamente con un interés paralelo por participar en estudios comparativos internacionales de evaluación del rendimiento educativo. Aparte de la valiosa experiencia regional del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, en el que participaron un total de trece países de América Latina y el Caribe, los estudios promovidos por organizaciones tan reconocidas como la IEA o la OCDE no han alcanzado todavía en la actualidad cotas altas de participación (Tiana, 2000). Durante los años '90, fueron más bien los programas de cooperación orientados hacia el reforzamiento de los sistemas nacionales de evaluación, como el impulsado en esa dirección por la OEI, los que lograron una participación más amplia.

Así pues, y hablando siempre en términos generales, la situación de la evaluación educativa en la región latinoamericana se caracteriza hasta el momento por la existencia de un contraste llamativo entre el desarrollo de ambiciosos programas nacionales de evaluación y una limitada participación en estudios comparativos internacionales. El desarrollo de esos programas nacionales está basado en la puesta en marcha de un buen número de instituciones y agencias de evaluación, cuya experiencia comienza ya a ser importante.

Una mirada a los sistemas nacionales de evaluación

Dando un paso más en el análisis, conviene que nos preguntemos por las principales características que presentan esos sistemas nacionales de evaluación que tanto auge adquirieron durante la última década. Tres son las preguntas clave que debemos hacernos, que están ligadas respectivamente a otras tantas dimensiones de la evaluación: ¿para qué se evalúa?, ¿qué se evalúa? y ¿qué uso se hace de la información?

a) Una multiplicidad de propósitos

La primera pregunta tiene que ver con los propósitos que orientan la evaluación de la educación. Y anticipando la respuesta, hay que reconocer que los sistemas nacionales de evaluación persiguen una multiplicidad de propósitos, al orientarse hacia la consecución de uno o varios de los siguientes objetivos:

- Ayudar a las escuelas y a los profesores a mejorar la calidad de la educación que imparten, proporcionándoles una información relevante y significativa que les permita valorar su situación real y sus logros.
- Conocer y valorar en qué medida se están alcanzando los objetivos educativos o los estándares propuestos para los distintos niveles o

parcelas del sistema educativo, contribuyendo al mismo tiempo a establecer metas y niveles de referencia.

- Forzar a las instituciones escolares y a los agentes educativos a generar dinámicas de cambio, por medio de la comparación o la competencia.
- Rendir cuentas a la sociedad o a las autoridades educativas acerca de los logros conseguidos, eliminando así la opacidad del sistema educativo y de las instituciones escolares.
- Establecer sistemas de incentivos para las instituciones y para los profesionales de la educación.
- Certificar, acreditar o seleccionar instituciones, programas de estudio, profesores o estudiantes.
- Conocer y valorar la situación del sistema educativo, así como el impacto producido por las políticas adoptadas.
- Introducir elementos de racionalidad en el debate público y en la construcción de políticas en el ámbito de la educación.

Es cierto que estos propósitos no son en todos los casos compatibles entre sí, pero no lo es menos que las políticas nacionales de evaluación seleccionan algunos frente a otros, les conceden distinto énfasis, establecen prioridades entre ellos o los interpretan de diferente manera. Así, por ejemplo, un propósito compartido por muchos, como es el caso de la incentivación del cambio a través de la comparación, no deja de ser objeto de polémica. Mientras que en algunos países ha llevado, por ejemplo, a la publicación de listas de clasificación de escuelas o universidades, con el ánimo de forzar una competencia abierta por conseguir más estudiantes o más recursos, en otros se ha rechazado frontalmente esa posibilidad y se han establecido mecanismos de confidencialidad de los datos que impiden ese tipo de uso, aunque permitiendo beneficiarse de la comparación. Como puede apreciarse, un mismo propósito puede inspirar políticas muy diferentes.

b) Una diversidad de actuaciones

La segunda pregunta tiene que ver con la actividad desarrollada por los sistemas nacionales de evaluación. Y también en este caso hay que anticipar la existencia de una considerable diversidad de modalidades de actuación, referidas a su vez a distintos ámbitos educativos. Entre los principales aspectos que son objeto de evaluación en la actualidad, hay que destacar cuatro campos prioritarios:

- Un primer campo que es objeto de considerable atención se refiere a la evaluación de los logros conseguidos por los alumnos. Muchas de las actividades desarrolladas en la actualidad por los sistemas nacionales de evaluación, y sin duda las más importantes de entre ellas, están orientadas

a la valoración de los resultados de la educación, medidos a través de los logros que consiguen los estudiantes. En la mayoría de los casos, esas actuaciones se centran en algunas áreas curriculares seleccionadas, especialmente las que se consideran fundamentales para la formación de los jóvenes (Lenguaje, Matemáticas, Ciencias, Historia). A veces se complementan con la evaluación de los logros conseguidos en otros ámbitos de la formación escolar, como la educación cívica, el dominio de las tecnologías de la información y la comunicación o las estrategias de aprendizaje. Las evaluaciones suelen estar referidas a etapas o momentos clave del proceso educativo, tales como el tránsito de una etapa a otra o el final de la escolarización obligatoria o de un determinado nivel de estudios. En ocasiones se evalúa a la totalidad de los estudiantes de los grados seleccionados, mientras que otras veces se trabaja solamente con muestras representativas. En cualquier caso, es ésta una línea de actuación prioritaria para los sistemas nacionales de evaluación.

- Un segundo campo que también atrae considerable atención se refiere al rendimiento que logran las instituciones educativas. El interés que ha generado la apertura de la "caja negra" que constituyen las instituciones ha estimulado el desarrollo de diversos programas orientados a su evaluación. Tanto las escuelas primarias y secundarias como las universidades han sido el objetivo prioritario de muchos planes de evaluación y de acreditación puestos en marcha en diversos países latinoamericanos. Esos planes de evaluación tienen características heterogéneas, pero suelen coincidir en conceder una atención especial a los resultados conseguidos por los estudiantes, aunque sin dejar generalmente de lado el análisis de la organización y el funcionamiento de la institución de que se trata.
- Un tercer campo que ha ido desarrollándose recientemente, si bien con mayores dificultades que los dos anteriores, es el relativo al desempeño profesional de los docentes. La convicción de que la actuación de los profesores está en la base de los resultados conseguidos por los estudiantes, aunque no sea su único determinante, ha generado un conjunto de actuaciones de diverso tipo, entre las que también se cuentan algunos programas de evaluación. Se trata de acciones que no han dejado de generar cierta polémica, tanto por sus características técnicas como por el tipo de uso que se hace de la información obtenida. En muchos casos, quizás en la mayoría, se ha optado por utilizar la evaluación docente en el marco de planes de incentivos profesionales, eludiendo sus aspectos más punitivos, pero en otros casos la evaluación se ha utilizado como instrumento de selección o de promoción profesional, o incluso se ha asociado con mejoras salariales o laborales.

- Un cuarto campo que también ha atraído el interés de los sistemas de evaluación se refiere al impacto de las políticas educativas adoptadas. En un contexto de reformas educativas, como las impulsadas en la región durante los '90, no es extraño que se haya planteado abiertamente la pregunta acerca de los efectos y el impacto de los procesos de cambio iniciados y de las medidas adoptadas. Desde esta perspectiva, han coexistido dos tendencias que deben diferenciarse, aunque ninguna de ellas haya llegado a atraer tanta atención como las arriba mencionadas. La primera se refiere a la evaluación de los programas más emblemáticos de los procesos de reforma, los denominados "programas estrella", aquellos en que las autoridades educativas han realizado una mayor inversión y depositado mayores esperanzas. La segunda se refiere a la evaluación de los propios procesos de reforma, con la pretensión de establecer un mecanismo de seguimiento sistemático de sus efectos. Pese a la limitada experiencia desarrollada en ambas direcciones, no cabe duda de que se trata de un campo llamado a experimentar un mayor desarrollo en el futuro próximo.

Aunque los cuatro campos mencionados no agotan la totalidad de los ámbitos abarcados por los sistemas nacionales de evaluación, no cabe duda de que se trata de los más destacados y los que han experimentado un mayor empuje. Nuevamente hay que insistir en que no todos ellos han sido desarrollados en todos los países, ni del mismo modo, ni con la misma energía, pero sí puede afirmarse que son objeto de atención especial por parte de los sistemas nacionales de evaluación.

c) Una pluralidad de usos

La tercera pregunta tiene que ver con el tipo de uso que se hace de la información obtenida por medio de la evaluación. Y al igual que en los dos casos anteriores, también hay que reconocer la existencia de una pluralidad de situaciones. Intentando analizar los usos más habituales de los resultados de la evaluación, podemos distinguir cinco tendencias predominantes:

- Una primera tendencia consiste, por paradójico que pueda resultar, en la renuncia a hacer uso de la información obtenida. Aunque no resulte muy justificable, ni desde el punto de vista de la eficacia administrativa, ni desde la perspectiva de la transparencia, ni menos aún desde las exigencias del comportamiento político democrático, hay que reconocer que existen casos en que los resultados de la evaluación no son publicados, ni difundidos, ni apenas utilizados.
- Una segunda tendencia, que es la más habitual, consiste en la elaboración y difusión de informes de evaluación. Los informes pueden ser de diverso tipo, tener una cobertura variable (nacional, regional, sectorial) y estar dirigidos a audiencias distintas (familias, profesores, autoridades edu-

cativas, investigadores y técnicos, etc.), llegando incluso a adoptar formatos diferentes. Lo importante en este caso no son las características concretas de los informes, que pueden ser de hecho muy diferentes, sino la confianza subyacente en que la difusión pública de los resultados puede satisfacer las demandas de información planteadas por los diversos agentes educativos. En algunos casos se han diseñado y aplicado estrategias bastante completas de difusión, logrando un impacto social y mediático importante.

- Una tercera tendencia consiste en la devolución confidencial de resultados a las instituciones participantes en la evaluación, con el propósito de que puedan utilizar los datos recibidos para elaborar sus propios planes de mejora, sin verse necesariamente expuestas al escrutinio público. Los modos concretos que puede adoptar esa devolución son muy variables, pero lo que caracteriza esa modalidad de actuación es su confianza en el valor intrínseco que encierra la comparación y en la capacidad de reflexión y de mejora que tienen las instituciones y los agentes de la educación cuando se enfrentan con su propia realidad.
- Una cuarta tendencia, no excesivamente habitual y desde luego controvertida, consiste en la publicación de los resultados obtenidos, en forma de tablas de clasificación. En algunos casos la publicación se realiza por los responsables de la evaluación, mientras que en otras ocasiones se permite o incluso se favorece que sean otros agentes (prensa, investigadores) quienes publiquen la información. Este tipo de uso resulta más amenazador para las instituciones educativas que los anteriores, motivo por el cual ha recibido críticas y generado resistencias.
- Una quinta tendencia consiste en la utilización de los resultados de la evaluación para proporcionar orientación pedagógica a instituciones y profesores. Este tipo de uso de la información obtenida a través de la evaluación adopta dos modalidades distintas. En primer lugar, son varios los sistemas nacionales que han elaborado instrumentos tales como cuadernillos de asesoramiento a los profesores, en los que se analizan las respuestas correctas de las pruebas, así como los errores más habituales que suelen producir los estudiantes, y se finaliza ofreciendo sugerencias didácticas más o menos concretas. En segundo lugar, aunque sea de forma más esporádica, también se han desarrollado acciones de orientación para instituciones determinadas, ayudándolas a elaborar sus propios planes de mejora, y se han incluido análisis de los resultados de la evaluación en programas de formación docente. Este tipo de uso, que es menos tradicional que la difusión de informes, está recibiendo una atención creciente por par-

te de diversas administraciones educativas.

Igual que sucedía en los epígrafes anteriores, también hay que decir aquí que los usos mencionados no son necesariamente excluyentes, sino que son cada vez más las administraciones educativas que los combinan en distintos modos. Es cierto que algunos de ellos resultan contrapuestos, como puede ser la alternativa existente entre la publicación de los resultados de las escuelas y la entrega confidencial de los datos, pero no siempre es el caso. De hecho, la combinación de la publicación de informes, la devolución de resultados a las instituciones participantes en la evaluación y la elaboración de instrumentos de orientación didáctica constituye una realidad bastante extendida en los países de la región.

A la vista de la situación que se acaba de exponer cabe extraer una primera conclusión, consistente en que los sistemas nacionales de evaluación de la región latinoamericana no recurren a estrategias de acción uniformes ni adoptan enfoques de evaluación idénticos. Y esta conclusión no es sólo una confirmación de los hechos que se han presentado, sino que se debe incluso considerar lógico que exista tal variedad.

En efecto, si aceptamos la existencia de una multiplicidad de propósitos, una diversidad de actuaciones y una pluralidad de usos, debemos aceptar la diversidad interna de los sistemas de evaluación. Serán precisamente los propósitos concretos que se persigan, los estilos de actuación que se adopten y los usos que se prefieran, los que determinarán las características del sistema de evaluación.

Una mirada a las pruebas nacionales de logros

Como se indicaba en el apartado anterior, los sistemas nacionales de evaluación que se han desarrollado en los países latinoamericanos desde los años '90 vienen prestando una atención especial a la medición de los logros conseguidos por los alumnos. Los instrumentos que se aplican con esa finalidad son las genéricamente denominadas pruebas nacionales, que merecen una mirada especial por la importancia que han llegado a adquirir.

Las pruebas nacionales presentan bastantes diferencias de unos países a otros, aunque también ofrecen algunas similitudes. Entre sus elementos comunes, quizás el más destacado sea la preocupación prácticamente universal que demuestran por evaluar el rendimiento alcanzado en Lenguaje y Matemáticas. No obstante, aunque se trata de dos áreas especialmente atendidas por los sistemas de evaluación, no hay que pensar que sean los únicos dominios evaluados. En efecto, la medición de los logros en ambas áreas suele ir acompañada por la evaluación de otras áreas curriculares complementarias (entre las que sobresalen las Ciencias, la Historia o las Lenguas extranjeras) o de otras capacidades de tipo transversal, no ligadas necesariamente a áreas especí-

ficas (como el autoconcepto, algunas estrategias de aprendizaje o ciertas actitudes).

Las pruebas nacionales suelen aplicarse preferentemente en los grados terminales de las principales etapas educativas (sobre todo, al final de la enseñanza primaria y de la secundaria). No obstante, también se aplican a veces en otros grados que tienen un especial interés desde el punto de vista diagnóstico (por ejemplo, el final de ciertos ciclos), aunque la identificación de tales puntos focales varíe de unos sistemas educativos a otros. El reciente desarrollo de algunos estudios internacionales ligados a la edad de los evaluados y no al grado que cursan, como es el caso del proyecto PISA de la OCDE, aún no ha encontrado traducción directa en los sistemas nacionales de evaluación latinoamericanos, ni es previsible que lo haga a corto plazo. El interés que despierta la valoración de los logros conseguidos al final de determinadas etapas supera hoy en la región al que suscita la valoración del nivel de formación a ciertas edades. La participación en aquellos proyectos internacionales parece satisfacer suficientemente las preguntas que puedan plantearse en este último sentido.

Dependiendo del propósito que guía a la evaluación, las pruebas se aplican en ocasiones a poblaciones o cohortes estudiantiles completas, mientras que en otros casos se limitan a muestras representativas de estudiantes. La primera decisión permite hacer un uso más singularizado de los datos obtenidos, tal como comparar el rendimiento de diversas escuelas o estudiantes, mientras que la segunda es más adecuada para usos diagnósticos. Por otra parte, hay países en que las pruebas se aplican cada año, aunque es más frecuente que se adopten ciclos plurianuales, de distinta duración.

Las pruebas nacionales parecen satisfacer ampliamente a las autoridades educativas, a la vista del uso tan extenso que se hace de ellas. No obstante, en un análisis más profundo también presentan insuficiencias y problemas, que requieren una consideración más atenta.

Tres son las principales insuficiencias de las pruebas nacionales de logros que han subrayado los especialistas (Ravela, 2000). La primera consiste en la existencia de algunas debilidades técnicas en los procesos de desarrollo y validación de los instrumentos de medición. La segunda se refiere a la insuficiente calidad y capacidad que las pruebas ofrecen para evaluar aprendizajes complejos. La tercera tiene que ver con el insuficiente aprovechamiento que suele hacerse de la información obtenida.

Los problemas a los que se ha hecho alusión son de distinto tipo. Para comenzar, habría que señalar algunos relativos al diseño y confección de los instrumentos de evaluación:

- Un primer problema tiene que ver con el contenido de las pruebas, asunto que plantea dos tipos de dificultades. La primera se refiere a la relación que debe existir entre las pruebas y los objetivos educativos. Lo deseable es que las pruebas estén referidas a los objetivos establecidos en el sistema educativo, de manera que permitan valorar el grado en que se alcanzan las metas previstas. Aunque este planteamiento resulta lógico y

plausible, plantea dificultades cuando no existen objetivos explícitos (lo que sucede en muchos casos), o cuando los objetivos de cada curso o etapa se formulan de manera poco concreta. Cuando no se pueden determinar los objetivos de forma precisa, resulta sin duda más difícil elaborar las pruebas. La segunda dificultad se refiere al tratamiento de la diversidad curricular existente entre diferentes regiones, distritos o escuelas. La mayoría de los modelos curriculares aplicados en América Latina conceden cierto grado de autonomía a las autoridades regionales o distritales o a las propias escuelas para definir el currículo. En consecuencia, no resulta sencillo hacer pruebas que abarquen todas las enseñanzas efectivamente impartidas en todo el sistema educativo. Las soluciones posibles son dos: elaborar pruebas de mínimos, lo que implica reducir el campo de evaluación, o elaborar pruebas comunes a partir de un consenso profesional o técnico. Ambas soluciones son utilizadas en la región, sin que quepa conceder prioridad a una de ellas sobre la otra.

- Un segundo problema tiene que ver con algunas disyuntivas que se plantean en relación con el diseño y la confección de las pruebas. La primera disyuntiva se refiere al enfoque de las pruebas, que pueden estar centradas en los conocimientos adquiridos por los alumnos o abarcar también el desarrollo de actitudes. Las pruebas que miden unos y otras no son iguales, ni siquiera plantean las mismas exigencias técnicas, lo que obliga a decidir su enfoque. Una segunda disyuntiva se refiere al carácter de las pruebas, pudiéndose optar por la elaboración de pruebas referidas a criterios de logro o por la confección de pruebas de tipo normativo, adaptadas a una distribución de resultados estadísticamente normal. La tercera disyuntiva se refiere a la delimitación de la población sujeta a evaluación, que permite optar por pruebas de aplicación censal (dirigidas a toda la población estudiantil de un determinado grado o edad) o de tipo muestral (dirigidas a muestras representativas de población).

Otro grupo importante de problemas están relacionados con la explicación de los resultados conseguidos. La mayoría de los sistemas nacionales de evaluación no suelen contentarse con medir los logros alcanzados por los estudiantes, sino que también pretenden explicar por qué se producen tales resultados. Sin embargo, dicho intento de explicación tropieza con varios problemas y dificultades:

- Un primer problema tiene que ver con el grado de coherencia realmente existente entre la evaluación y el currículo.
- Como se señalaba al comienzo del trabajo, el desarrollo de los sistemas nacionales de eva-

luación ha ido generalmente asociado a otros procesos de reforma curricular.

- No obstante, esa simultaneidad no ha asegurado siempre la existencia de una adecuada coherencia entre ambos aspectos. En los casos más extremos, puede incluso decirse que ambos procesos se han desarrollado en direcciones divergentes. Cuanta más importancia se ha concedido a las pruebas, mayor ha sido su impacto sobre el desarrollo curricular.
- Un segundo problema, menos complejo pero no menos influyente, está relacionado con el tipo de interpretación que permiten los diversos modelos de pruebas. Como se señalaba más arriba, dos son los modelos de pruebas utilizados más habitualmente: criteriosales y normativas. Las pruebas de tipo criterial proporcionan una base sólida para valorar la suficiencia de los logros conseguidos, ya que se apoyan en una definición operativa de los objetivos que deben alcanzarse. Por el contrario, las pruebas de tipo normativo exigen alguna operación adicional para poder valorar la suficiencia de los logros alcanzados, dado que se refieren a situaciones de normalidad estadística. La existencia de estos dos modelos alternativos obliga a considerar cuidadosamente qué usos se quieren hacer de las pruebas, antes de decidir el que conviene aplicar.
- Un tercer problema se refiere a las dificultades conceptuales y técnicas que se plantean a la hora de buscar explicaciones. Por una parte, hay que reconocer la debilidad teórica de los modelos habituales de producción educativa, que en buena parte deriva del problema que plantea la causalidad en educación. Esa debilidad influye en la selección de los factores que se consideran asociados al rendimiento, pocos de los cuales cuentan con confirmación empírica sólida (Scheerens, 1996). Además, la explicación de los resultados se ha basado tradicionalmente en la distinción de dos tipos de variables, unas extrínsecas y otras intrínsecas. Mientras que las primeras (nivel socioeconómico y cultural, recursos destinados a la educación) han sido objeto de análisis abundantes, las segundas (organización del sistema educativo, procesos institucionales, procesos de aula) vienen resultando menos concluyentes y aún requieren investigaciones adicionales.
- Un cuarto problema tiene que ver con la voluntad que a veces existe de buscar muchas vías de explicación de los resultados obtenidos, que puede traducirse en un exceso de datos recogidos. Cuando ese exceso de datos va unido, como ocurre en ocasiones, con una insuficiente delimitación conceptual de las variables seleccionadas y con limitaciones en el análisis estadístico de los datos, la situación puede llegar a resultar inmanejable. La consecuencia suele ser una evidente

infrautilización de los datos y una ineficiencia en el uso de los recursos disponibles.

- Un último grupo de problemas que plantean las pruebas de logros están relacionados con el uso que se hace de las mismas y con las interpretaciones a que dan pie:
 - Un primer problema tiene que ver con los distintos tipos de uso que pueden hacerse de la información. Los resultados de la evaluación pueden utilizarse con una finalidad formativa, orientada al desarrollo institucional y profesional de los agentes implicados, para fomentar la competición, e incluso para impulsar políticas de mercado en el ámbito educativo. Obviamente, las interpretaciones que permiten tales posibilidades son diferentes y persiguen distintos efectos. Es necesario señalar que, si bien esos son los extremos de un continuo de posibilidades de uso, existen otras opciones intermedias en que dichos propósitos pueden combinarse en diferente proporción. La decisión sobre el tipo de uso que se pretende hacer de los datos de la evaluación condiciona en buena medida el diseño de la evaluación.
 - Un segundo problema tiene que ver con la existencia de diversas posibilidades de comparación (con una norma o criterio, con otras realidades semejantes, consigo mismo a lo largo del tiempo), cada una de las cuales plantea sus propias exigencias y ofrece diferentes posibilidades. El uso de la comparación plantea un problema complicado, que tiene que ver con la justicia de la comparación. El intento de dar respuesta a esa cuestión ha impulsado el desarrollo de técnicas de cálculo del denominado valor añadido.
 - Un tercer problema consiste en el excesivo reduccionismo en que a veces se cae cuando se pretende interpretar los resultados de la evaluación. Ese defecto puede venir originado por una excesiva limitación del contenido de las pruebas, que no permite llegar a conclusiones razonables sobre el logro de los objetivos propuestos, o por una interpretación poco rigurosa de los datos, que puede incluso llegar a distorsionar las interpretaciones.

A la vista del análisis que acaba de hacerse de las pruebas nacionales de logros, se llega a la conclusión de que constituyen la práctica de evaluación más habitual en la región, encontrándose bastante asentadas en la actualidad. No obstante, como se ha puesto de manifiesto, presentan algunas deficiencias y problemas que deberán solventarse en el futuro, si se quiere asegurar su credibilidad y se pretende que contribuyan a mejorar la educación.

Algunas reflexiones finales

Si bien es cierto que las pruebas nacionales de logros constituyen la práctica más habitual de evaluación educativa en América Latina, no es la única que se desarrolla en la actualidad. Como se indicaba en las primeras páginas del trabajo, existen otros tres ámbitos que son también objeto de evaluación. El primero corresponde a los programas y las políticas de educación, que despiertan especial interés en el actual contexto de revisión de las reformas educativas puestas en marcha durante los años '90. En sus formas más elaboradas ha servido para diseñar mecanismos de seguimiento de los procesos de cambio, aunque se trata de una práctica escasamente implantada. El segundo ámbito corresponde a las instituciones educativas, para cuya evaluación se han diseñado y aplicado algunos modelos, tradicionales o innovadores, aunque sin que pueda decirse que esta tendencia haya alcanzado todavía suficiente desarrollo. El avance más importante que se ha logrado en este sentido consiste en la expansión de una conciencia nueva acerca del valor que tiene la evaluación institucional con vistas a la mejora de la calidad de la educación. El tercer ámbito, y al mismo tiempo el más controvertido, corresponde al desempeño docente. Aunque en algunos países se han desarrollado mecanismos de valoración de la tarea docente que están conectados con la medición del rendimiento de los estudiantes, no existen todavía experiencias suficientes ni de larga tradición en este sentido.

Las experiencias que han explorado otras direcciones son aún más escasas.

En conjunto, puede afirmarse que estos otros tres ámbitos abarcados por la evaluación educativa están actualmente infra-desarrollados en relación con las pruebas de rendimiento. Hay que subrayar, no obstante, que la evaluación de tales aspectos puede contribuir notablemente a la mejora de la calidad de la educación. Precisamente de esa potencialidad procede el interés que despiertan en muchos países de la región.

Como conclusión final, cabe afirmar que la evaluación es una realidad integrada que cuenta con varios componentes: estudios internacionales de evaluación, estudios diagnósticos nacionales del rendimiento, evaluación de las instituciones educativas y otras evaluaciones complementarias de diversos aspectos de la actividad educativa (preparación y desempeño de los docentes, impacto de las políticas aplicadas, funcionamiento de algunos programas relevantes). Aunque no todos ellos se han desarrollado con la misma energía, los sistemas nacionales de evaluación no deberían infravalorar la importancia de concebirlos de forma integrada, de manera que las acciones emprendidas se refuercen mutuamente.

El colofón a estas reflexiones que provoca la mirada a los sistemas nacionales de evaluación educativa consistiría en que la evaluación debe considerarse como un instrumento para la mejora de la calidad de la educación y no como una panacea capaz de solucionar todos los problemas que la educación plantea. Otra consideración diferente de ésta no puede sino conducir a algunos errores de cierta importancia.

Bibliografía:

- *Braslavsky, C. (2001): Educación secundaria: ¿cambio o inmutabilidad?, Buenos Aires, Santillana.*
- *Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe (2001): Quedándonos atrás. Un informe del progreso educativo en América Latina, Santiago de Chile, PREAL.*
- *Gajardo, M. (1999): Reformas educativas en América Latina. Balance de una década, Santiago de Chile, PREAL (documento de trabajo n° 15).*
- *Ravela, P., ed. (2000): Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?, Santiago de Chile, PREAL.*
- *Ravela, P. (2001): ¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?, Santiago de Chile, PREAL.*
- *Scheerens, J. (1996): "Can the School Effectiveness Knowledge Base Guide School Management?" en II International Conference on School Management. Participative Management and School Evaluation, Bilbao, Universidad de Deusto, pp. 98-119.*
- *Tiana, A. (1996): "La evaluación de los sistemas educativos", Revista Iberoamericana de Educación, n° 10, 1996, pp. 37-61.*
- *Tiana, A. (2000): Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación, Washington, BID – Departamento de Desarrollo Sostenible.*
- *Winkler, D. y Gershberg, A.I. (2000): Los efectos de la descentralización del sistema educacional sobre la calidad de la educación en América Latina, Santiago de Chile, PREAL (documento de trabajo, n° 17).*

“LAS EVALUACIONES EDUCATIVAS QUE AMÉRICA LATINA NECESITA”

Presentación

Este documento aborda la importancia, propósitos y usos de las evaluaciones estandarizadas a gran escala de aprendizajes y/o logros educativos en América Latina y el Caribe, básicamente en los niveles primario y medio. Está dirigido a quienes formulan políticas educativas, docentes, académicos, empresarios, sindicatos, organizaciones sociales, agencias de financiamiento y medios de prensa, con el fin de aportar al debate y decisiones sobre la evaluación estandarizada en los sistemas educativos.

Se entiende por evaluación estandarizada en gran escala a aquella que permite producir información comparable acerca de los desempeños de estudiantes pertenecientes a distintos contextos culturales y regionales e, incluso, a distintos países, y que ofrece un panorama de la situación de un país o de un estado o provincia, aun cuando ello se haga a veces a través de una muestra no demasiado grande (por ejemplo, 5.000 estudiantes).

El documento se centra en evaluaciones de aprendizaje –definido como el cambio ocurrido en los conocimientos y capacidades de cada estudiante a lo largo del año lectivo y/o de logro educativo, entendido como la acumulación de conocimientos y capacidades a lo largo de toda la vida del estudiante. No se pretende orientar aquí acerca de otros aspectos también importantes de la evaluación del quehacer educativo, como son las evaluaciones del desempeño docente, de las políticas educativas, de los centros educativos o aquellas que realizan los docentes en las aulas, entre otras. Tampoco se pretende incursionar en la evaluación en el nivel terciario ni en las pruebas de selección para el mismo.

Los contenidos de esta publicación ayudarán a quienes toman decisiones de política educativa a comprender y analizar las diversas opciones existentes en cuanto a los propósitos y usos de los sistemas de evaluación y las implicancias de cada una de ellas, de modo de orientarlos en cómo concebir una estrategia de evaluación.

Es preciso evitar el simplismo y la ingenuidad con que muchos piensan en la evaluación, lo que genera sistemas mal concebidos y deficientemente implementados y, como consecuencia de ello, efectos perversos para el sistema educativo, malgasto de recursos y descrédito de la evaluación externa entre los docentes. Por el contrario, se debe reconocer la gran variedad de aspectos que deben considerarse al momento de implementar un sistema de evaluación o reformar uno existente. Para que la inversión en evaluación valga la pena, debe tener propósitos claros, una filosofía orientada a construir una visión de responsabilidad compartida en relación a la educación, un diseño técnico de calidad y adecuado

a los propósitos, una orientación fuerte a apoyar a los docentes en su tarea y una voluntad política manifiesta de encarar acciones dirigidas a resolver los problemas y deficiencias que la evaluación ponga de manifiesto.

Asimismo, requiere de una inversión importante para construir una unidad técnica competente y un plan de largo plazo, lo cual lleva tiempo y no puede ser improvisado

I. ¿Por qué son importantes las evaluaciones nacionales de logros educativos?

La realización de evaluaciones estandarizadas como forma de conocer mejor la dinámica de procesos y resultados en los sistemas educativos es cada vez más frecuente a nivel regional y mundial, en países de muy diversas culturas y orientaciones ideológicas de gobierno. Prueba de ello es la creciente participación de los países en las evaluaciones internacionales como PISA, TIMSS y PIRLS, y regionales como SERCE (en América Latina) y SACMEQ (en África), así como el desarrollo de diferentes tipos de sistemas nacionales y subnacionales de evaluación.

En algunos casos, la evaluación está motivada por una preocupación por la formación ciudadana y la consolidación de una sociedad democrática; en otros, por la productividad de la fuerza de trabajo y la competitividad de la economía nacional, por las oportunidades para el desarrollo integral de las personas y sus posibilidades de participación en la sociedad del conocimiento, o por la equidad y la visión de la educación como uno de los caminos para superar la pobreza. Muchos sistemas de evaluación parten de una combinación de los intereses anteriores. En casi todos los casos se asume que la evaluación puede servir:

- como base para adoptar políticas educativas mejor fundamentadas;
- para mejorar la gestión de los sistemas educativos;
- como instrumento para la colaboración y el aprendizaje continuo al interior de los mismos.

A continuación se destacan los aportes principales de este tipo de evaluaciones.

La evaluación estandarizada ayuda a visualizar los resultados educativos del conjunto de estudiantes

La educación es una actividad “opaca”, en el sentido de que sus resultados no son directa ni inmediatamente observables. Un buen docente puede apreciar si sus estudiantes están aprendiendo o no y cómo, pero no todos los docentes tienen los mismos criterios de valoración, los cuales están fuertemente vinculados a su experiencia profesional: su formación general y específica, su conocimiento de la disciplina que enseña, su capacidad para percibir los procesos y dificultades de los estudiantes, su familiarización con diversos tipos de alumnos, etc. Dada la heterogeneidad del cuerpo docente de un país, no es posible obtener una visión de la situación del conjunto por la mera agregación de los puntos de vista individuales. La evaluación estandarizada apunta a lograrlo.

Las evaluaciones entregan información sobre el real acceso al conocimiento y a las capacidades que alcanzan los alumnos, más allá de la cantidad de años de estudio que tengan

En décadas pasadas, la equivalencia entre permanencia en el sistema educativo y acceso al conocimiento y a las capacidades simbólicas se daba por descontada y los indicadores para valorar a los sistemas educativos estaban relacionados con el acceso (matrícula, cobertura, retención, etc.). En esos tiempos, las mayorías más pobres y con menor capital cultural no accedían al sistema educativo o apenas cursaban unos pocos años de educación básica. Con la progresiva universalización del acceso al sistema educativo, hoy muchos estudiantes procedentes de los sectores sociales más desfavorecidos llegan en desventaja en términos de manejo del lenguaje oral (oficial) y escrito, con lo cual la equivalencia entre años de estudio y acceso al conocimiento y a las capacidades ya no es tal. Por otra parte, el crecimiento del acceso a la profesión docente no fue acompañado por garantías de calidad en su formación, lo cual también cuestiona la mencionada equivalencia.

Así, más años en el sistema educativo no necesariamente significa que todos los niños y jóvenes estén incorporando los conocimientos, actitudes y capacidades necesarias para la vida personal y social, los que son cada vez más sofisticados y complejos. Las evaluaciones buscan dar luces sobre lo que está ocurriendo al respecto.

Las evaluaciones estandarizadas ayudan a hacer visibles un conjunto de aspectos centrales de la labor educativa

Entre otras cosas, las evaluaciones estandarizadas aportan información sobre:

- en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados o niveles;

- cuál es el grado de equidad o inequidad en el logro de dichos aprendizajes;
- cómo evolucionan a lo largo de los años, tanto los niveles de logro como la equidad en el acceso al conocimiento por parte de los diversos grupos sociales;
- en qué medida y cómo las desigualdades sociales y culturales inciden sobre las oportunidades de aprendizajes de los estudiantes;
- cuál es la diversidad de prácticas educativas existentes en escuelas y maestros y cómo las mismas se relacionan con los aprendizajes de los estudiantes en diversos contextos sociales;
- cómo influyen las condiciones de la enseñanza (situación de los maestros, recursos disponibles, tiempos de estudio, etc.) en los progresos de los alumnos;
- qué efecto tienen en los logros educativos las inversiones en programas educativos, los cambios en la estructura del sistema, los cambios curriculares, los programas de formación, la adquisición de material educativo, etc.

Un sistema de evaluación de aprendizajes y/o logros educativos puede aportar información importante a diversos actores sociales

En la medida que el sistema produzca y comunique adecuadamente información sobre los aspectos antes indicados, puede ser un instrumento clave de mejora, enriqueciendo la comprensión de la situación educativa y la toma de decisiones en diversos ámbitos:

- Las autoridades y los encargados de formular políticas educativas pueden comprender mejor los problemas de la enseñanza y del aprendizaje; hacerse cargo de las carencias en que se desarrolla la labor docente y desarrollar políticas pertinentes para apoyar el trabajo de las escuelas. Las evaluaciones también les permiten valorar, sobre una base de evidencia empírica sólida, el impacto de las políticas y programas que han impulsado y los probables efectos de las que se proponen impulsar.
- Los directivos y docentes pueden, a partir de una mirada externa sobre los logros educativos en el conjunto del sistema, comprender mejor lo que están logrando y lo que no están logrando sus propios estudiantes, cómo están aprendiendo y qué dificultades tienen. Pueden aprender de las experiencias de otros docentes y escuelas que trabajan con estudiantes de características tanto similares como diferentes a los propios. Pueden tomar decisiones más apropiadas acerca de qué

aspectos del currículo enfatizar y enriquecer sus propios modos de evaluar el aprendizaje de sus estudiantes.

- Los supervisores y encargados de la formación de docentes pueden apoyarse en la información sobre los logros y dificultades educativas en el conjunto del sistema para estudiar en profundidad las debilidades en el enfoque o en la práctica de la enseñanza que probablemente generan algunas de las insuficiencias constatadas en los aprendizajes. Desde esa nueva perspectiva, pueden mejorar su labor tanto de orientación a los docentes como de formación de los mismos. En particular, los supervisores pueden beneficiarse de contar, como instrumento de apoyo para su labor, con un mapa de las escuelas que las caracterice tanto en términos de la composición sociocultural de su alumnado como de sus logros educativos.
- Los padres y madres de estudiantes, adecuadamente informados, pueden comprender mejor qué se espera que sus hijos aprendan, qué es lo que están logrando y qué pueden hacer para colaborar con la escuela y con el aprendizaje de sus hijos.
- La ciudadanía en general estará mejor informada sobre lo que acontece al interior del sistema educativo y, por tanto, estará más atenta a los temas y problemas de la educación. Estará, además, en mejores condiciones para exigir, tanto a los poderes públicos como a los profesionales de la docencia, la mejora continua de la educación que se brinda a los niños, niñas y jóvenes, y el uso responsable de los recursos que se destinan a la educación.

El desarrollo de un sistema nacional de evaluación estandarizada puede obligar a realizar una discusión informada sobre qué aspectos del currículo formal son exigibles a todos los estudiantes y a definir con claridad qué es lo que todos los estudiantes deberían haber aprendido al final de cada ciclo educativo.

La mayoría de los currículos en la región están constituidos por largas listas de objetivos y temas, todos deseables pero no todos realizables. La elaboración de pruebas nacionales obliga a definir qué es lo que debe ser considerado como fundamental y, por tanto, lo que todos los estudiantes deben saber y ser capaces de hacer. Se pueden usar distintos términos para designar a estas definiciones: estándares, competencias fundamentales, indicadores de logro, niveles de desempeño, metas de aprendizaje, criterios de suficiencia, entre otros.

Para no confundir...

Al desarrollar estos sistemas nacionales de evaluación, es necesario explicitar tanto el rol que estos cumplen así como las funciones que no cumplen y cuáles son sus limitaciones, para evitar así los riesgos de mala interpretación respecto a lo que son y la información que recogen y difunden.

– Una evaluación estandarizada aporta información fundamental e indispensable sobre la “calidad educativa”, aunque no sea un indicador completo de la misma.

No todos los objetivos valiosos de la educación están incluidos en este tipo de evaluaciones. Hay una gran cantidad de saberes, actitudes, valores y aprendizajes relevantes que no pueden –por la dificultad de medirlos de esta manera– o no deben –porque son propios de cada entorno local y, por tanto, no son exigibles a todos los estudiantes del país– formar parte de una evaluación estandarizada.

– La evaluación estandarizada de aprendizajes y/o logros educativos es un componente esencial de un sistema integral de evaluación educativa, pero no es el único tipo de evaluación relevante.

También tienen importancia la evaluación realizada en el aula por los docentes, la evaluación del desempeño docente, la evaluación de los centros educativos, la evaluación de las políticas educativas, la evaluación del uso de los recursos, la evaluación de la relevancia del currículo, etc.

– La evaluación es condición necesaria, pero no suficiente, para mejorar la educación.

Si bien existe alguna evidencia de que la mera existencia y difusión de información tiene algún impacto sobre ciertos actores, es necesario recordar que la evaluación es apenas uno de varios elementos clave de la política educativa que debe procurarse estén adecuadamente alineados: la formación docente inicial y en servicio, la carrera y condiciones para el trabajo docente, la gestión escolar, la supervisión, los diseños curriculares, los libros y materiales educativos, una inversión de recursos proporcional a las necesidades de las diversas poblaciones y una acción decidida por parte de los responsables educativos para resolver los problemas detectados, entre otros.

– La evaluación estandarizada externa solo tendrá efectos positivos sobre la educación si es concebida, percibida y empleada como un mecanismo de responsabilización pública de todos los actores vinculados al quehacer educativo.

Existe siempre el riesgo de que la política educativa se concentre en la implementación de evaluaciones, pero que luego no se tomen acciones concretas para enfrentar y resolver los problemas que las mismas ponen de manifiesto. Muchas veces las autoridades se limitan a informar de los resultados y transferir toda la responsabilidad por la solución de los problemas detectados a las escuelas y familias. Otras veces, toda la responsabilidad se atribuye a los docentes, quienes, por su parte, tienden a transferir toda la responsabilidad a los padres o al contexto. Lo importante es evitar ingresar en una dinámica de culpabilización y, por el contrario, intentar construir una lógica de responsabilidad compartida en torno a la educación.

II. La situación de la evaluación estandarizada en América Latina

Los sistemas nacionales de evaluación estandarizada se desarrollaron con fuerza en toda la región durante los años 90. Algunos países han mantenido sus sistemas en funcionamiento con continuidad, aun cuando se produjeran cambios de enfoque o variaciones en la institucionalidad de los mismos. Otros países han tenido importantes discontinuidades y han tenido que comenzar prácticamente desde cero en repetidas oportunidades o tendrán que hacerlo en el futuro próximo.

Además de ello, actualmente:

- Dieciséis países están participando en el Segundo Estudio Regional en 3° y 6° de primaria que lleva adelante la OREALC/UNESCO: Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.
- Seis de estos países participan en PISA 2006 y otros tres se incorporarían al ciclo PISA 2009.
- Algunos países de la región han participado además (o lo están haciendo) en los estudios de Matemática y Ciencias (TIMSS), Lectura (PIRLS) y Educación Cívica que lleva adelante la International Association for the Evaluation of Educational Achievement (IEA).
- Durante los últimos 12 años ha estado activa en la región la red de los sistemas de evaluación organizada en torno al Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de OREALC/ UNESCO, en la que participan actualmente casi todos los países latinoamericanos. Esta red se reúne dos veces al año y es un espacio de intercambio de experiencias y formación en el ámbito de la evaluación.
- Más recientemente se ha constituido el Grupo

Iberoamericano de PISA, integrado por Argentina, Brasil, Chile, Colombia, España, México, Portugal y Uruguay.

Las evaluaciones a gran escala en la región no solo son más frecuentes sino también están mejorando

Los principales cambios y mejoras que se observan en los últimos años incluyen:

- Mayor transparencia en la difusión de los resultados. Parece haberse superado la etapa en que las autoridades en varios países interferían en la divulgación de los resultados de las evaluaciones cuando estos no les parecían favorables.
- Tendencia creciente a pasar de pruebas normativas –que tienen como propósito principal ordenar comparativamente a los estudiantes – a pruebas de criterios, que se enfocan en qué es lo que los estudiantes saben y son capaces de hacer. Asimismo, crecientemente se incluye en las pruebas de criterios una definición de cuál es el resultado que todos los estudiantes deberían alcanzar para que su desempeño pueda ser considerado satisfactorio.
- Mejoría de las capacidades técnico-metodológicas para la construcción de pruebas y para el procesamiento de los datos. Existen esfuerzos por desarrollar pruebas que evalúen un rango de conocimientos y capacidades más amplio y que incluyan preguntas de respuesta construida. Se han ido incorporando también metodologías más sofisticadas para el tratamiento de datos, como la Teoría de Respuesta al Ítem y el Análisis Multinivel.
- Creciente atención a la difusión y uso de los resultados. Se ha ido comprendiendo que no basta con implementar una evaluación y publicar un informe, sino que es necesario desarrollar una estrategia de divulgación y un conjunto de reportes apropiados a cada una de las audiencias a las que se quiere llegar. Las Unidades de Evaluación tienen una mayor conciencia acerca de los usos apropiados e inapropiados para cada tipo de evaluación, aunque esto muchas veces no está suficientemente claro entre quienes toman decisiones de política.
- Mayor preocupación por la investigación de los factores que inciden sobre los aprendizajes. Si bien falta mucho por hacer en términos metodológicos e interpretativos para producir buenos trabajos de investigación, en la mayoría de los países hay conciencia de que es preciso avanzar en la investigación y en la construcción de hipótesis sobre cómo las dinámicas de la gestión de los sistemas educativos, los procesos escolares,

las prácticas de enseñanza y las decisiones de política educativa influyen en las oportunidades de aprendizaje de los niños y niñas.

- Creciente participación de los países en evaluaciones internacionales. Esto ha tenido consecuencias positivas en las Unidades de Evaluación y en la calidad de su trabajo: ha contribuido a la construcción y acumulación de capacidad técnica en evaluación, ha facilitado el intercambio y el desarrollo de un lenguaje común entre países y ha ayudado a mejorar la calidad de diversos procesos técnicos (elaboración de pruebas, incorporación de preguntas de respuesta abierta, muestreo, control de las aplicaciones, análisis de datos, modos de difusión, entre otros).

Las principales debilidades presentes en los sistemas de evaluación estandarizada en la región se observan en la estrategia de difusión, en el uso de los resultados y en la calidad técnica de las evaluaciones.

A pesar de las mejoras previamente mencionadas, persiste un conjunto importante de debilidades que es preciso abordar:

- Es frecuente que desde el nivel político se demande la implementación de grandes operativos de evaluación sin propósitos claramente definidos y en tiempos y con recursos absolutamente insuficientes para hacerlo técnicamente bien.
- Aún no se tiene claro que no cualquier evaluación sirve para cualquier propósito y que es necesario diseñar cuidadosamente el sistema de evaluación para el largo plazo, en función de propósitos y usos claramente establecidos. Esto es determinante para las decisiones relativas a la utilización de censos o muestras, a los grados y disciplinas que serán evaluados, la periodicidad de las evaluaciones, el tipo de pruebas y escalas de reporte, entre otros.
- Falta una mayor discusión pública sobre qué debe ser evaluado y qué aspectos del currículo deben haber sido enseñados a todos los estudiantes al concluir determinados niveles educativos. Como resultado de ello, las metas y/o estándares de aprendizaje no son claros. Mientras no exista claridad al respecto, tampoco existirá claridad ni para la enseñanza ni para la evaluación. Esta tarea requiere de un amplio debate social y de un trabajo técnico coordinado entre las unidades de currículo y de evaluación.
- Para mejorar la investigación sobre los factores que explican los resultados, falta mejorar mucho los instrumentos complementarios que se aplican junto con las pruebas y concebir diseños de investigación más ambiciosos: estudios longitu-

dinales, estudios de “valor agregado” y estudios cualitativos, entre otros.

- Los Ministerios de Educación han tenido, por lo general, una limitada capacidad para concebir, formular e implementar políticas que den respuesta a los problemas identificados en las evaluaciones. Falta para ello mayor articulación entre las Unidades de Evaluación y otros actores educativos relevantes, dentro y fuera de los Ministerios de Educación, y mayor articulación entre la evaluación, el desarrollo curricular, la formación inicial y el desarrollo profesional de los docentes.
- Las acciones para la divulgación y uso de los resultados de las evaluaciones aún son insuficientes, en particular para lograr que los docentes los comprendan y utilicen y para que los resultados se incorporen a la cultura escolar.
- En la mayoría de los países perdura la práctica de divulgar resultados de tipos de escuela, escuelas individuales, estados o provincias, sin una adecuada consideración de los contextos socioculturales en que operan las instituciones y subsistemas educativos. Esto conduce a interpretaciones y conclusiones erróneas sobre la eficacia educativa de dichas instituciones o subsistemas. Para ello es necesario realizar evaluaciones de “valor agregado”, que implican dos mediciones sobre una misma población en distintos momentos. Sobre este aspecto la experiencia en la región es mínima.
- En muchos países persisten importantes debilidades técnicas en el diseño de las evaluaciones: excesiva simplicidad de la mayor parte de las preguntas y dificultad para elaborar preguntas que permitan evaluar capacidades cognitivas complejas; focalización de las evaluaciones en los logros educativos correspondientes a un grado específico, lo que impide saber qué han aprendido quienes aún no logran lo esperado para dicho grado; deficiencias en la conformación de las muestras y en la forma de estimar y reportar los errores de medición; y debilidades para establecer mediciones que sean comparables en el tiempo.
- No hay suficientes profesionales calificados para diseñar y conducir este tipo de evaluaciones, lo que se agrava por la falta de continuidad de los equipos técnicos de los países, muchas veces por razones de índole político-partidaria. Esto dificulta la acumulación de conocimiento y experiencia en la región y ha dado lugar a que muchos países tengan que comenzar desde “cero” con sus sistemas de evaluación, algunos años después de haber tenido uno en funcionamiento y haberlo desmantelado.

El riesgo de las malas evaluaciones

Solo una evaluación técnicamente buena y cuyos resultados sean conocidos y utilizados de manera apropiada, puede tener algún impacto en la mejora de los aprendizajes. Como es obvio, una evaluación cuyos resultados son poco conocidos y utilizados supone un ejercicio estéril y un desperdicio de recursos.

Más graves aún son aquellas situaciones en que evaluaciones técnicamente malas son ampliamente difundidas o en que las evaluaciones son empleadas para fines diferentes de aquellos para los cuales fueron concebidas y diseñadas, lo cual puede tener efectos dañinos para el sistema educativo. El razonamiento según el cual lo que importa es evaluar y difundir resultados como mecanismo de rendición de cuentas, sin importar la calidad de la evaluación, es una falacia que puede hacer retroceder los sistemas de evaluación e imposibilitar una discusión seria de las ventajas y desventajas de la rendición de cuentas o responsabilización por los resultados.

III. Los propósitos y los usos de los resultados de las evaluaciones estandarizadas

Sea en la implementación de un sistema de evaluación o en su reforma, es necesario considerar algunas opciones básicas de acuerdo a lo que se espera de él.

El primer paso es definir cuál será el propósito del sistema de evaluación y para qué se usarán sus resultados

Es preciso considerar diversas opciones, que no son necesariamente excluyentes. Un sistema puede combinar varias de ellas, pero es importante tener conciencia de que cada opción requiere de un determinado diseño y tiene exigencias técnicas y de costos diferentes. Se debe analizar si el sistema nacional de evaluación será:

- de carácter diagnóstico (para alumnos, escuelas o sistema educativo), sin consecuencias directas para los estudiantes, cuya finalidad principal es enriquecer con información de calidad las percepciones, decisiones y acciones de diversos actores del sistema –autoridades y cuerpos técnicos, supervisores, directivos, docentes, estudiantes y sus familias – para mejorar la enseñanza y el aprendizaje.
- de certificación de los logros educativos de los estudiantes, cuyo propósito central es establecer quiénes han alcanzado los conocimientos y desempeños necesarios para aprobar un curso o nivel y, en consecuencia, aprueba o reprueba.

Es necesario también establecer un adecuado balance entre la implementación de pruebas nacionales y la

participación en pruebas internacionales

Las pruebas nacionales pueden ofrecer una mirada más apropiada acerca de qué aprenden los estudiantes de aquello que se les está enseñando, mientras que las pruebas internacionales pueden servir para ubicar la situación del país en relación a lo que saben y son capaces de hacer los estudiantes en otras sociedades y enriquecer el debate sobre el currículo nacional y el enfoque de la enseñanza. Los países deberían analizar cuidadosamente en qué pruebas internacionales participar, teniendo en cuenta lo que cada una pretende evaluar y su pertinencia en relación a los objetivos nacionales. Parece recomendable participar periódicamente en al menos una evaluación regional o internacional, en el marco de la política de prioridades que cada país defina.

Evaluaciones para la certificación de los aprendizajes de los estudiantes

Un sistema de evaluación para certificar los aprendizajes de los estudiantes a través de exámenes nacionales de alta calidad técnica tiene ventajas asociadas a la transparencia y responsabilidad por los resultados.

Hoy es común que dos estudiantes que han aprobado un mismo nivel educativo en dos regiones distintas de un país, posean en realidad niveles de conocimiento muy diferentes. Los sistemas de evaluación para la certificación dan transparencia al valor de los certificados educativos ante la sociedad. Además, dichos sistemas hacen responsables tanto a los docentes como a los propios estudiantes por alcanzar los conocimientos y capacidades que evalúa el examen, lo cual tiene impactos positivos sobre el aprendizaje. En todo caso, la evaluación externa con propósitos de certificación es más apropiada para los ciclos superiores del sistema educativo, sobre todo hacia el final de la educación media.

Un sistema de evaluación para la certificación implica algunos requerimientos que son costosos

Por una parte, las pruebas deben ser de carácter censal y cada estudiante debe tener más de una oportunidad para rendirla, lo cual requiere implementar varios operativos de evaluación cada año. Por otra parte, las pruebas deben tener una amplia cobertura curricular, lo cual requiere que sean extensas, en lo posible con preguntas de respuesta construida (que tienen un importante costo de codificación) y deben abarcar varias asignaturas o disciplinas.

Un sistema de evaluación para la certificación puede generar tensiones importantes que es necesario anticipar

Si las pruebas son exigentes, ello puede conducir a niveles de reprobación muy altos, que afectarían principalmente a los sectores sociales más vulnerables y podrían

agudizar problemas como la deserción y el desempleo juvenil, lo cual las haría insostenibles desde el punto de vista social y político. Por ello, es preciso acompañar este tipo de evaluaciones (en realidad todo tipo de evaluaciones, pero estas en particular) con acciones pedagógicas de carácter complementario y compensatorio. Proponer estándares altos implica asumir responsabilidad sistémica por proporcionar todas las oportunidades para aprender: textos, materiales, instalaciones y preparación de maestros, entre otros. Existen diversos caminos para enfrentar este problema:

- Establecer que la evaluación estandarizada constituya solo una parte del proceso de certificación de aprendizajes de los estudiantes (por ejemplo, el 40% de su calificación final). El resto de la nota correspondería a las evaluaciones que realizan los docentes a cargo de los estudiantes. Si bien en este caso se mantiene el problema respecto a la heterogeneidad de los criterios de evaluación empleados por los docentes, es una forma de combinar la evaluación externa y la interna, e incluso de aportar a los docentes una mirada externa que les permita reflexionar sobre sus propios criterios.
- Establecer un período de transición hacia estándares únicos y universalmente aplicables. Durante dicho período, el foco principal se centra en la mejoría o progreso de los estudiantes de cada escuela respecto a su situación anterior, y no únicamente en el cumplimiento del estándar absoluto de certificación.

La peor respuesta que podría darse a este problema sería establecer un porcentaje fijo de reprobación, por ejemplo, no debe reprobarse más del 10% de los estudiantes. Ello implica utilizar pruebas más fáciles y, por lo tanto, enviar a estudiantes, familias y educadores una señal equivocada respecto a las expectativas de aprendizaje.

EVALUACIONES DE CARÁCTER DIAGNÓSTICO Y FORMATIVO

Un sistema de evaluación de tipo formativo y sin consecuencias directas para los estudiantes tiene ventajas en términos de costos, posibilidades de diseño y de establecimiento de estándares altos, y aporte a una cultura de evaluación

Los costos de aplicación de estas pruebas pueden ser menores, dado que pueden hacerse en base a muestras y aplicarse solo en algunos grados claves y cada cierto número de años. Por otra parte, las pruebas pueden diseñarse con carácter matricial, en que no todos los estudiantes responden a las mismas preguntas, sino a bloques de las mismas, lo cual permite trabajar con una cantidad muy grande de preguntas y obtener un análisis más detallado de los distintos aspectos del currículo. Estos sistemas permiten definir estándares o niveles de

expectativa exigentes sin que ello produzca un fracaso generalizado. Asimismo, ayudan a construir una cultura de la evaluación y a acumular capacidad técnica, de modo que, cuando se analice la posibilidad de establecer un sistema con consecuencias, existan las condiciones para hacerlo apropiadamente. Una de las alternativas a considerar entre los usos de carácter formativo es la de “liberar” una prueba completa para que sea aplicada en forma autónoma por parte de los maestros, con el fin de ayudarles a identificar dificultades de alumnos individuales y de enriquecer su repertorio de instrumentos de evaluación.

El principal problema de estas evaluaciones es que pueden carecer de impacto si no se toman algunas acciones complementarias

Las evaluaciones de carácter diagnóstico y formativo pueden resultar intrascendentes si no van acompañadas de una estrategia precisa y de una inversión importante para asegurar la divulgación de los resultados y su uso en acciones educativas posteriores, puesto que su efectividad se juega en que los diferentes actores reciban, comprendan y utilicen los resultados.

Por esto hay que tener presente que:

- Para que los resultados de este tipo de evaluaciones tengan impacto sobre las políticas educativas, se debe invertir tiempo en el análisis y discusión de los mismos por parte de diversas áreas del Ministerio de Educación y otros actores relevantes, en la comprensión de los problemas y deficiencias que los resultados ponen de manifiesto y en la concepción de acciones e inversiones apropiadas para hacerles frente. Las autoridades deben estar dispuestas a someter sus políticas y decisiones al escrutinio de la ciudadanía, para lo cual es necesario también invertir en la comunicación apropiada y permanente de los resultados a la opinión pública.
- Para que los resultados tengan impacto sobre las prácticas de enseñanza, es imprescindible asimismo invertir tiempo en el análisis e interpretación de sus implicancias didácticas: si los estudiantes no son capaces de resolver cierto tipo de situaciones, ¿qué es lo que se está haciendo de manera insuficiente o inapropiada en las aulas y qué es lo que se debería hacer? Este tipo de análisis debe ser realizado tanto por especialistas en la didáctica de las áreas evaluadas como por los docentes, creando para este último efecto espacios permanentes y sistemáticos de formación en servicio y de trabajo colectivo al interior de las escuelas. Es muy importante que los maestros puedan analizar la mayor cantidad posible de ítems para poder identificar cuáles revelan un bloqueo importante para el desarrollo de nuevos conceptos o capacidades. Sin embargo, siempre será necesario conservar el carácter confidencial

de una parte de los ítemes para poder realizar mediciones comparables en el tiempo.

- Para que los resultados tengan impacto sobre la motivación de los estudiantes y sobre la actitud de las familias hacia el aprendizaje escolar, es necesario proveerles de información apropiada y comprensible acerca de lo que se considera imprescindible haber aprendido en cada grado o nivel educativo y de las acciones que pueden contribuir al logro de dichos aprendizajes.

Evaluar con una perspectiva amplia

Es muy importante que las pruebas no se limiten a evaluar los conocimientos y competencias correspondientes a un único grado (aquel en que se aplica la evaluación), sino que incluyan una perspectiva más amplia de niveles de desempeño –desde más básicos a más complejos – a lo largo de varios grados, de modo de identificar qué han aprendido los estudiantes en grados anteriores y qué necesitan ahora. De este modo, los docentes podrán constatar y subsanar las deficiencias en los aprendizajes de cursos previos, que impiden a los estudiantes continuar avanzando y los resultados serán útiles no solo para los docentes del grado evaluado, sino también para los de grados anteriores. El énfasis no estará puesto en comunicar a los docentes que los estudiantes “aprueban” o “desaprueban”, sino en comunicar que están en diferentes puntos de un continuo de aprendizaje, en el que todos pueden y necesitan progresar.

Las pruebas de tipo diagnóstico pueden tener carácter muestral o censal, dependiendo de la estrategia de cambio educativo

El carácter censal o muestral tiene diversas implicancias, aunque también es posible combinar una aplicación por muestras controlada con una distribución censal de las pruebas para su aplicación en forma autónoma por parte de las escuelas, con fines formativos y de análisis de resultados e identificación de estudiantes que necesitan apoyos complementarios.

- Las pruebas a base de muestras sirven para tener un diagnóstico global del sistema. Requieren de un cuidadoso diseño de la muestra, de modo de obtener información representativa para los niveles de desagregación en los que se desee actuar y tomar decisiones (regional, provincial, municipal; urbano y rural; escuelas indígenas, etc.). El impacto de los resultados de estas pruebas depende principalmente de las medidas de política educativa que se tomen a nivel central y de una estrategia de difusión apropiada que llegue a todas las escuelas.
- Las pruebas censales sirven para obtener información de cada una de las escuelas e incluso de los alumnos. El impacto de sus resultados depende de la devolución de la información a cada

comunidad educativa, con un enfoque y un formato adecuados para promover una mayor participación y compromiso a nivel local. La información también puede resultar muy útil para dirigir mejor las políticas hacia distritos o escuelas con mayores problemas, ya que se puede contar con un “mapa” de resultados de todas las escuelas, zonas, provincias, tipos de escuela, etc.

Uso de las evaluaciones para establecer incentivos

Además de las opciones examinadas, existen políticas de evaluación estandarizadas dirigidas, a establecer incentivos económicos en función de los resultados o a propiciar un mercado competitivo entre las escuelas

Tres son las principales modalidades en este tipo de políticas:

- Utilizar los resultados para construir rankings de escuelas y hacerlos públicos, como forma de: fomentar la responsabilidad de las escuelas por sus resultados; entregar a las familias información para la toma de decisiones sobre la escuela a la cual enviar a sus hijos; promover la competencia entre los centros educativos por obtener mejores resultados.
- Utilizar los resultados para entregar incentivos económicos a las escuelas que obtienen mejores resultados o mejoras en relación a sus resultados en evaluaciones anteriores.
- Utilizar los resultados como indicador de la calidad del trabajo de cada docente y como criterio para otorgar incentivos económicos.

Estos enfoques de uso de la evaluación a veces se constituyen, con o sin intención deliberada, en un mecanismo por el cual el Estado renuncia a su responsabilidad por los resultados del sistema educativo

En ocasiones, el Estado limita su accionar a la realización de evaluaciones, la entrega de resultados y el establecimiento de incentivos en función de estos resultados, transfiriendo la responsabilidad por los resultados a la relación entre escuelas y familias, como si se tratase de un asunto entre actores privados, en lugar de ocuparse de crear las condiciones para que la enseñanza resulte efectiva proveyendo los recursos necesarios, constituyendo un elenco docente adecuadamente formado y estableciendo mecanismos idóneos para la evaluación y la orientación de la labor de las escuelas. Este modo de operar no considera la complejidad de la labor educativa, en especial en medios socialmente desfavorecidos, y la necesidad de invertir en la creación de capacidades como herramienta principal para la mejora de la enseñanza y el aprendizaje.

Crterios básicos para la divulgación y uso de los resultados

La responsabilidad por los resultados debe ser compartida por diversos actores

Debería evitarse todo uso de los resultados con el fin deliberado o implícito de culpabilizar o responsabilizar de manera exclusiva a ciertos actores. La responsabilidad debe ser compartida entre las autoridades nacionales y regionales, los cuerpos docentes, los estudiantes y sus familias.

La responsabilidad por los resultados requiere de una adecuada combinación entre las exigencias y los apoyos a las escuelas y a los docentes

Por un lado, es necesario que escuelas y docentes asuman su responsabilidad respecto al objetivo de que todos los estudiantes aprendan lo que se espera de ellos. Simultáneamente, es deber de las autoridades establecer las políticas de apoyo necesarias para que escuelas y docentes puedan desarrollar su labor de manera apropiada. Un esquema de exigencias altas sobre escuelas y docentes sin los correspondientes apoyos solo puede generar malestar y desánimo. Un esquema de apoyos sin contrapartes y exigencias puede generar autocomplacencia.

Es inapropiado utilizar los resultados de las pruebas estandarizadas como indicador principal de la calidad del trabajo del docente o de la escuela

Esto es importante, en particular si no se ha controlado el efecto de otros factores internos y externos al sistema educativo y si no se tiene en cuenta que el aprendizaje depende también de la motivación y esfuerzo personal del estudiante. Además, dado que los resultados de pruebas estandarizadas no son el único indicador de la calidad de la educación, es preciso considerarlos en conjunto con otros aspectos relevantes como la retención, la pertinencia y relevancia de lo que se enseña, la formación de valores y hábitos o la formación ciudadana. Debe evitarse la identificación del término "calidad" con resultados de pruebas estandarizadas.

Para facilitar la comprensión y uso de los resultados, no basta con ofrecer resultados numéricos

Es fundamental ilustrar a los distintos actores acerca del tipo de tareas que los estudiantes debieron resolver en las pruebas. Este criterio debe combinarse adecuadamente con la necesidad de mantener el carácter confidencial de un conjunto de preguntas que puedan emplearse en futuras evaluaciones para establecer mediciones comparables en el tiempo. Si bien no se pretende que los docentes utilicen pruebas estandarizadas

para evaluar a sus estudiantes, el conocimiento y la comprensión de las mismas pueden contribuir a mejorar sus modos de evaluar y a desarrollar una cultura de la evaluación.

Al reportar las diferencias de resultados en función de la composición sociocultural del alumnado de las escuelas, se debe cuidar de no crear un sistema de expectativas de logro diferenciado por grupos sociales

En todo caso, las diferencias deberían producirse en las políticas para crear condiciones apropiadas para enseñar a los grupos más desfavorecidos. Se debería descartar también cualquier tipo de uso de los resultados que, directa o indirectamente, propicie la selección de estudiantes por parte de las escuelas a los efectos de mejorar sus resultados.

Advertencias sobre la comparación de resultados entre escuelas

Aunque los resultados de pruebas estandarizadas no dan cuenta en forma exhaustiva de la calidad educativa de una escuela, sí aportan información relevante de los niveles de desempeño alcanzados

La información sobre resultados en dos pruebas estandarizadas – en general, lenguaje y matemática – no constituye por sí misma una evaluación de la "calidad educativa" de las escuelas ni debería presentarse al público como tal. La "calidad" de una escuela incluye otros aspectos relevantes y valorados por los docentes, estudiantes y familias, tales como el desarrollo emocional, las relaciones interpersonales, la formación cívica y en valores, entre otros. Aun así, la información comparativa acerca de los niveles de desempeño alcanzados por los estudiantes en una diversidad de escuelas puede ser relevante para los equipos docentes de las escuelas, en la medida en que enriquece la percepción de la propia labor, permitiéndoles ubicar los logros de sus estudiantes en el contexto de los alcanzados por los estudiantes en otras escuelas.

Para que la comparación entre los niveles de desempeño de los estudiantes sea apropiada, se deben considerar sus condiciones sociales de origen

Los resultados académicos de las escuelas deberían poder compararse con los de escuelas de similar composición social, dado que los retos y dificultades que implica enseñar a estudiantes de origen desfavorecido (o de lengua materna indígena) son muy diferentes que en el caso de estudiantes de familias con educación secundaria completa y/o terciaria. También hay que tener en cuenta las tasas de deserción de las escuelas y sus políticas de selección de estudiantes, dado que una escuela puede mejorar sus resultados a través de la exclusión de estudiantes con dificultades.

Las comparaciones entre escuelas deben considerar la diferencia entre medir el “logro educativo” y medir el “aprendizaje”

En rigor, si el propósito es entregar a las escuelas, familias y/o autoridades, información acerca de la capacidad de enseñar de los docentes y las escuelas, es necesario medir el progreso de los estudiantes a lo largo de un período lectivo (“aprendizaje”) y no únicamente el resultado final (“logro”). Las diferencias entre estos dos términos se expresan en que:

- El aprendizaje puede ser definido como el cambio ocurrido en los conocimientos y capacidades de cada estudiante a lo largo del año lectivo. Su medición requiere la aplicación de dos pruebas, una al inicio y otra al final. De este modo es posible constatar el progreso realizado por cada alumno.
- El “logro educativo” en cambio, se mide con una única prueba y refleja la acumulación de conocimientos y capacidades a lo largo de toda la vida del estudiante, lo cual incluye el capital cultural familiar y sus experiencias en otras escuelas o con otros docentes.

Debe tenerse siempre presente que el aprendizaje y el logro educativo de los estudiantes dependen tanto de lo que hacen las escuelas y docentes, como del esfuerzo de los propios estudiantes, del apoyo de las familias a la tarea escolar, del contexto comunitario y cultural y de las políticas educativas. Los problemas educativos no pueden ser resueltos apelando exclusiva ni principalmente a mecanismos de mercado.

Los resultados expresados en términos de rankings de escuelas deben ser mirados con precaución

La mayoría de los rankings de escuelas transmiten una falsa imagen de ordenamiento. Una escuela puede ocupar el primer lugar y otra el número 40 y, sin embargo, las diferencias entre sus promedios pueden no ser estadísticamente significativos. Esto significa que no puede decirse si un promedio es realmente superior al otro, porque las diferencias entran dentro de los márgenes de error de la medición. Pero aun cuando las diferencias en los promedios sean estadísticamente significativas, pueden ser irrelevantes en términos de las proporciones de alumnos que logran los niveles de desempeño esperados.

Itinerario para la toma de decisiones

Considerando las alternativas existentes para una política de evaluación, antes de poner en marcha un sistema de evaluación (o de modificar uno existente) es indispensable seguir un itinerario de decisiones para definir las características del sistema que incluya al menos

las siguientes preguntas básicas:

- ¿Para qué se quiere evaluar? ¿Quiénes usarán los resultados y con qué propósitos? ¿Qué se espera saber que no se sepa ya?
- ¿Cuáles serán las unidades de análisis para el reporte de resultados: estudiantes individuales, grupo de clase/maestros, escuelas, tipos de escuela, entidades subnacionales, sistema educativo?
- De acuerdo a los propósitos definidos, ¿qué es más apropiado evaluar – aprendizaje o logro educativo – al egreso de determinados grados o ciclos?
- ¿Qué tipo de consecuencias tendrán los resultados y para quiénes?
- De acuerdo a los propósitos definidos, ¿es necesario trabajar en forma censal o son suficientes evaluaciones en base a muestras?
- ¿Qué grados y qué disciplinas es importante evaluar?
- ¿Cada cuánto tiempo es necesario y adecuado realizar las evaluaciones?

Las respuestas a estas preguntas deberían plasmarse en un plan de evaluación a corto, mediano y largo plazo, claro y explícito. En la elaboración de dicho plan es muy importante considerar los costos en dinero implicados en cada opción y los recursos humanos necesarios para su adecuada implementación. En particular, se debe balancear la inversión a realizar en la recogida de información con la inversión en la difusión y uso de los resultados. Muchos Ministerios de Educación destinan importantes sumas de dinero a recoger una gran cantidad de información todos los años, que luego casi no se analiza, ni se difunde ni se utiliza. Es absolutamente inconducente implementar un primer operativo de evaluación si no se ha elaborado un plan de trabajo de largo plazo.

“Mandatos” útiles al momento de tomar decisiones políticas respecto al sistema de evaluación

- **“No harás daño”**. Una de las primeras normas del juramento hipocrático señala que el médico se abstendrá de todo aquello que pueda generar daño o perjuicio a sus pacientes. Del mismo modo, a la hora de pensar en un sistema de evaluación es importante considerar los riesgos de efectos no deseados y perjudiciales que la estrategia de evaluación diseñada podría tener sobre el propio sistema educativo que se desea mejorar.
- **“El consumo excesivo de este producto puede resultar perjudicial para la salud”**. La eva-

luación no puede estar por encima de la educación. El exceso de evaluaciones puede producir perjuicios para la salud del sistema educativo, en especial si solo hay evaluaciones pero no hay políticas para dar respuesta a los problemas que estas ponen de manifiesto.

- **“No evaluarás en vano”.** Se debe evitar recoger todos los años información que nunca será analizada ni utilizada. Para que un sistema de evaluación tenga impacto, debe tener una periodicidad que garantice que los datos puedan ser analizados, discutidos, comprendidos y utilizados. Incorporar la nueva información y traducirla en decisiones y acciones lleva tiempo. Los cambios en el sistema educativo requieren más tiempo aún.
- **“Vísteme despacio, Sancho, que estoy apurado”.** Es muy importante que quienes toman las decisiones en los Ministerios de Educación desestimen la creencia ingenua en la existencia de caminos fáciles y rápidos. No es posible montar un programa de evaluación serio en tres meses. No existe coyuntura ni “ventana de oportunidad” que lo justifique: las consecuencias de la improvisación se harán sentir tarde o temprano. La evaluación requiere reflexión cuidadosa acerca de sus fines y usos; discusión pública sobre qué es lo fundamental a evaluar; diálogo e involucramiento de diferentes actores; conformación de equipos técnicos competentes en diversos aspectos; e información previa para que los actores se involucren con la evaluación.

IV. Los desafíos de calidad técnica de las evaluaciones

Una vez definida la política de evaluación, es necesario implementarla de acuerdo con diversos estándares apropiados de calidad técnica. Esto implica abordar los siguientes desafíos:

Elaborar un referente o marco conceptual que defina con precisión los conocimientos y desempeños que son considerados apropiados al finalizar el grado o ciclo educativo que será evaluado

Si bien esta construcción exige antes que nada un debate y decisiones de carácter político, las definiciones tomadas en dicho terreno deben luego ser adecuadamente traducidas en especificaciones y estándares de carácter técnico. Asimismo, el debate político-educativo debe estar alimentado por información sobre los avances conceptuales recientes en cuanto a la enseñanza y a los desempeños de los estudiantes en las disciplinas consideradas.

Incluir en las pruebas actividades con diversos grados de complejidad

Se debe contemplar actividades cuya solución requiera de competencias cognitivas suficientemente complejas, apropiadas a los desafíos de la sociedad del conocimiento, así como también actividades sencillas que permitan informar sobre el punto en que se encuentran los estudiantes menos avanzados. Las actividades deben poseer una serie de propiedades psicométricas que es necesario garantizar y cuyo control requiere pruebas piloto y análisis cuidadoso. También es importante, en la medida de lo posible, ampliar el uso de preguntas de respuesta construida, con el doble propósito de evaluar capacidades más complejas y de mejorar la articulación entre la evaluación externa y la cultura docente y escolar.

Conformar las pruebas a partir de la apropiada integración de las actividades en bloques y cuadernillos

Este es un desafío técnico importante y complejo, que requiere de conocimientos especializados y experiencia. Especial atención debe otorgarse a la decisión de utilizar la Teoría Clásica o la Teoría de la Respuesta al Ítem para la construcción y el análisis de las pruebas, lo cual requiere de programas de procesamiento modernos, analistas bien entrenados, análisis rigurosos y asesoramiento de alto nivel.

Definir los “puntos de corte” que establecen los límites entre niveles de desempeño en una prueba

Vinculado con esto, se debe definir la metodología a seguir para establecer cuál de los niveles debe ser considerado como aceptable para un estudiante al finalizar el grado o ciclo educativo evaluado. Un rendimiento aceptable no puede ser establecido automáticamente como equivalente al 51% o más del máximo posible en una prueba.

Diseñar muestras apropiadas a los propósitos de la evaluación

Esto tiene por objeto evitar operaciones más grandes y costosas de las estrictamente necesarias y que, al mismo tiempo, tengan un nivel de precisión adecuado. En relación a este último aspecto debe incluirse, además, la estimación y reporte de los márgenes de error de las mediciones.

Definir cómo se equiparán las evaluaciones

La equiparación de las evaluaciones se refiere a la metodología mediante la cual los resultados de pruebas aplicadas en distintos años se hacen comparables. Este es uno de los desafíos técnicos más importantes

que los sistemas de evaluación deben afrontar si uno de sus propósitos es informar sobre los avances o retrocesos en los logros educativos a lo largo del tiempo. Este proceso es indispensable para poder afirmar que los eventuales cambios que se constaten responden a cambios en la realidad educativa y no meramente a cambios en el instrumento de medición. Requiere tomar recaudos tanto desde el punto de vista estadístico, como desde el punto de vista de la continuidad de los conocimientos y competencias evaluadas, y la estabilidad en la longitud y la estructura de la prueba.

Construir datos longitudinales que permitan apreciar la evolución en el tiempo de los aprendizajes de un mismo conjunto de estudiantes

Este tipo de datos, que exigen más de una medición en el tiempo, es el que evalúa el "aprendizaje" entendido como cambio y el que mejor permite establecer cuáles son los procesos propiamente escolares que tienen influencia sobre el mismo. Este enfoque es el que permitiría obtener información más relevante para la política educativa y para la investigación.

Establecer mecanismos de control del cumplimiento de las condiciones estandarizadas que deben regir durante la aplicación de las pruebas

El cumplimiento de las condiciones de aplicación de las pruebas involucra aspectos técnicos complejos y muchas veces descuidados. Sin embargo, son fundamentales para que la información sea confiable y comparable. Esto apunta a temas como:

- La calidad de la capacitación de los aplicadores.
- El establecimiento de controles de calidad durante la aplicación de las pruebas.
- La motivación a los estudiantes para realizar las pruebas.
- La logística de la distribución.
- El retorno seguro de los materiales.

Combinar las evaluaciones con estudios cualitativos

Es importante combinar las evaluaciones nacionales estandarizadas, que ofrecen una mirada sobre el conjunto del sistema, con la realización de estudios cualitativos que permitan profundizar en los procesos escolares y los procesos que ocurren en las aulas. Esta combinación de enfoques es el mejor modo de aportar información rica y compleja para pensar en las políticas educativas y en las prácticas de enseñanza.

La necesaria transparencia

Sobre todos los procesos técnicos debe existir información transparente y accesible. Es necesario dar un especial énfasis a la documentación de los procedimientos técnicos seguidos en:

- La construcción de los instrumentos.
- La estimación de la precisión de las mediciones (y, por consiguiente, su margen de error).
- El diseño de las muestras y la cobertura alcanzada.
- La aplicación y control de calidad de la misma.
- La definición de niveles de desempeño y puntos de corte.
- La equiparación y comparabilidad de los resultados con mediciones anteriores.

V. Sobre la constitución de las unidades de evaluación

Para llevar adelante una política de evaluación de logros educativos o aprendizajes apropiada, es necesario contar con Unidades de Evaluación con las capacidades y los recursos adecuados que implica este complejo emprendimiento.

La puesta en marcha de un sistema de evaluación serio requiere de un tiempo mínimo de entre dos y tres años

Este periodo es necesario para llevar adelante los siguientes procesos fundamentales:

- Discusión, definición y difusión pública de los propósitos del sistema de evaluación, del tipo de consecuencias, de los usos esperados y de qué debe ser evaluado.
- Diseño del plan de evaluación a largo plazo.
- Conformación de equipos técnicos con la diversidad de capacidades requeridas (elaboración de pruebas; elaboración de cuestionarios; conocimiento de las disciplinas a evaluar y de su didáctica; currículo y estándares; muestreo; logística de la aplicación y control de calidad de la misma; captura y limpieza de datos; procesamiento y análisis; construcción de escalas; contextualización sociocultural de los resultados y análisis de factores asociados; interpretación de los datos en términos de políticas educativas, de didáctica y de prácticas de enseñanza).

- Elaboración de instrumentos, pilotaje y supervisión externa de los mismos.

Las Unidades de Evaluación necesitan tener una cierta estabilidad en el tiempo de sus elencos técnicos

El desarrollo de un sistema de evaluación requiere de una planificación a 10 ó 15 años. Si los equipos técnicos cambian con frecuencia, se desperdicia el conocimiento y la experiencia acumulada en un área compleja y se desacreditan los procesos de evaluación ante la sociedad y los educadores.

Las Unidades de Evaluación necesitan tener independencia para reportar los resultados de su trabajo

Al igual que en el caso de las Unidades de Estadísticas sociales y económicas, las unidades que realizan las acciones de evaluación y la divulgación de los correspondientes resultados no deben depender de los tiempos e intereses político-partidarios. Se ha discutido mucho respecto a si las Unidades de Evaluación deberían estar dentro o fuera de los Ministerios de Educación. El argumento principal para esta última opción es justamente la necesaria independencia y transparencia. Sin embargo, la experiencia en la región muestra casos de países con unidades estables e independientes dentro de los Ministerios de Educación y otros con unidades inestables que no han logrado consolidar su trabajo aun estando en una institución externa. La respuesta a esta disyuntiva es que no importa tanto el lugar institucional como la cultura de continuidad y transparencia que se cree en torno a la evaluación. Esto se logra cuando existe un mandato claro y una institucionalidad sólida en relación a la evaluación, lo cual implica algún tipo de estatuto jurídico para el sistema de evaluación. Establecer esto por ley es uno de los caminos a considerar, dado que exige un acuerdo amplio, en lo posible suprapartidario, que permita establecer un plan de evaluación de la educación de largo plazo. En este terreno, al igual que en el económico, se requiere de cierta estabilidad en las reglas de juego. Si la política de evaluación cambia permanentemente, se genera desconfianza y se pierde credibilidad.

Una institucionalidad sólida requiere de órganos de gobierno y de asesorías técnicas, independientes y plurales, y de un presupuesto apropiado y plazas de trabajo que garanticen la operación de la unidad con la calidad técnica requerida

La independencia de la Unidad de Evaluación no debe dar lugar a su desvinculación de la política educativa. Por el contrario, la evaluación debe responder a un proyecto político-educativo con amplio respaldo y debe mantener una estrecha relación con otras áreas claves de la política educativa como la formación docente, el desarrollo curricular, el planeamiento y formulación de proyectos, la evaluación de programas y la investigación.

Para que las evaluaciones estandarizadas tengan impacto sobre las políticas y las prácticas educativas, es necesario generar nuevas interfases y modalidades de trabajo que aseguren el alineamiento de los distintos actores y ámbitos de decisión

Las Unidades de Evaluación deben asumir que su tarea va más allá de producir datos. Si su misión es producir información para otros actores, se requiere de personas preparadas y con tiempo para establecer líneas de diálogo con instancias y actores responsables de:

- Análisis de política educativa, con el fin de mejorar tanto los diseños de las evaluaciones como los planes de procesamiento y los reportes de resultados, teniendo en cuenta preguntas relevantes para la política educativa (por ejemplo, la muestra puede ser diseñada de modo que permita evaluar el impacto de políticas o programas específicos en ciertos grupos de escuelas).
- Análisis didáctico, con el fin de interpretar los resultados y los problemas de aprendizaje de los estudiantes desde la perspectiva de la enseñanza y la didáctica de la disciplina evaluada, elaborar reportes con significado didáctico para los docentes, diseñar programas de formación en servicio a partir de los resultados, pensar la articulación de las evaluaciones estandarizadas con la evaluación en el aula.
- Comunicación, con el fin de elaborar una diversidad de reportes de resultados, con diversos formatos y lenguajes, que sean apropiados y comprensibles para diversas audiencias.

Evaluar bien requiere inversión

Más vale no hacer ninguna evaluación que hacer una evaluación mala o deficiente. Esto debe ser tenido en cuenta al tomar las decisiones acerca del Plan de Evaluación. Asimismo, más vale un sistema de evaluación modesto pero con costos asequibles y sostenibles en el tiempo, que una gran evaluación, amplia y sofisticada, que solo podrá realizarse una vez y que luego no podrá replicarse.

La inversión en evaluación debe ser valorada en función del uso que se hace de sus resultados, más que a partir de otros indicadores como el costo por alumno.

Los costos de las evaluaciones son relativamente bajos cuando se los compara con los presupuestos nacionales y con las inversiones alternativas. Pero cualquier inversión en evaluación, alta o baja, es inútil si no se hace uso de los resultados.

VI. Diez recomendaciones acerca de las evaluaciones que la región necesita

1. La evaluación debe concebirse como un elemento articulado en un conjunto más amplio de acciones y políticas educativas.

La evaluación por sí misma no produce mejoras. Deben existir articulaciones estables entre el área de evaluación y las áreas de desarrollo curricular, formación docente, investigación, diseño de políticas, comunicación y prensa, entre otros.

2. La evaluación debe contemplar un proceso de reflexión colectiva sobre el estado de la educación y los caminos para mejorarla.

El primer paso para ello es la consulta y el debate públicos acerca de qué deben aprender los estudiantes y acerca de los propósitos y consecuencias de la evaluación. Del mismo modo, es fundamental la discusión pública constructiva a partir de los resultados, con el fin de encarar las insuficiencias e inequidades en el acceso al conocimiento por parte de los y las estudiantes. Para ello es necesario invertir en comunicación y difusión tantos o más recursos que en la propia evaluación, antes, durante y después de la misma.

3. La evaluación debe estar al servicio del desarrollo de un sentido de responsabilidad compartida por la educación como bien público.

Debe promover el compromiso con la educación de todos los actores, cada uno según su lugar y ámbito de acción. Por este motivo, debe evitarse utilizar la evaluación para culpabilizar a actores específicos por los problemas detectados.

4. Los sistemas de evaluación de la región necesitan ampliar progresivamente el abanico de fines educativos que son objeto de evaluación.

Es importante en este sentido incluir a la formación ciudadana, otras disciplinas además de Lenguaje y Matemática, así como un espectro más amplio de competencias y capacidades.

5. Los sistemas de evaluación de la región deberían progresivamente diseñar evaluaciones de progreso de los estudiantes, dado que son las que pueden aportar mayor información acerca del impacto de las políticas educativas, de las acciones de las escuelas y de las prácticas de enseñanza, en los aprendizajes de los y las estudiantes.

6. Un sistema de evaluación es un proyecto de largo plazo, por lo que requiere de un compromiso del Estado y de una planificación cuidadosa del diseño del sistema.

Ello exige tomar decisiones acerca de los fines, las

consecuencias, los aspectos a evaluar, las áreas y los grados, y la periodicidad de las evaluaciones, entre otras cosas. Todo ello exige tiempo, por lo que no es conveniente pretender la implementación de sistemas de evaluación en plazos reducidos.

7. Un buen sistema de evaluación requiere inversión, principalmente en la conformación de equipos humanos calificados, así como recursos económicos suficientes para una adecuada implementación de todos los procesos implicados.

8. El sistema de evaluación debe sustentarse en una actitud firme de transparencia en relación a los resultados y de rendición de cuentas a la sociedad.

9. Los ministerios de educación deben asumir un compromiso serio y consistente con los resultados de la evaluación, lo cual implica promover el diálogo acerca de los problemas detectados y los modos de enfrentarlos, diseñar líneas de acción apropiadas para resolver los problemas e invertir los recursos necesarios para llevarlas adelante.

10. Los sistemas de evaluación deben ser objeto de evaluación periódica, con el fin de analizar la calidad técnica de la información que producen y su relevancia para diversos actores educativos y sociales.

“UNA MIRADA TÉCNICO-PEDAGÓGICA ACERCA DE LAS EVALUACIONES DE CALIDAD EDUCATIVA”

Introducción

Las preguntas sugeridas como punto de partida para la reflexión de esta mesa fueron tres: ¿Qué estamos evaluando? ¿Qué de lo evaluado resulta significativo? ¿Por qué las reformas educativas realizadas a lo largo de la década del 90 no se ven traducidas en los resultados de las evaluaciones?

Para aventurar la respuesta que considero debe darse a esas tres preguntas considero necesario presentar antes la que considero apropiada para una pregunta más: ¿qué deberíamos evaluar de modo que la evaluación sirva para mejorar la calidad educativa?

Organizaré mi presentación, pues, en cuatro apartados, cada uno de los cuales pretende responder una de las cuatro preguntas mencionadas.

¿Qué deberíamos evaluar para que la evaluación sirva para mejorar la calidad?

Ante un auditorio interesado en la educación no es necesario abundar en lo relativo a la importancia de su calidad y su evaluación. Pero teniendo en cuenta la diversidad de concepciones de estas dos nociones (calidad y evaluación) juzgo necesario explicitar mis propias ideas al respecto, de modo que las respuestas que ofreceré para las tres preguntas siguientes cobren sentido pleno.

Tratando, pues, de responder a esta primera pregunta, afirmo que, idealmente, una evaluación adecuada para que sirva realmente para mejorar la calidad educativa sería una muy amplia, que incluyera todas las dimensiones de la calidad; todas las áreas del currículo; los aspectos cognitivos pero también los actitudinales y valorales; no sólo niveles de dominio elementales, sino también los más elevados; insumos, pero también procesos y productos; los diversos actores y niveles organizacionales del sistema educativo; y que, además, incluyera el análisis de los factores que inciden en la calidad, de manera que dé bases para el diseño y la implementación de políticas.

Desarrollo el primero de los elementos anteriores: una buena evaluación deberá comprender todas las dimensiones de la calidad.

Sabemos que las políticas educativas han puesto el acento sucesivamente en la atención de la cobertura, luego en la eficiencia terminal y, más recientemente, en el nivel de aprendizaje alcanzado por los alumnos y en la equidad del servicio educativo. La coincidencia de la secuencia anterior de tales políticas en muchos países de

la región latinoamericana y otras partes del mundo no es casual: responde, desde luego, a la similar evolución del contexto demográfico, económico, social y cultural.

Es frecuente que el término calidad se defina en forma restringida, limitándolo a lo que se refiere a los niveles de aprendizaje. En este sentido se dice que, tras centrar la atención en la cobertura y la eficiencia, ahora las políticas educativas ponen el énfasis en la calidad. Otra manera de usar la palabra, que juzgo preferible, le da un sentido más amplio, que se aplica también a las políticas de cobertura y eficiencia terminal.

Un concepto amplio de calidad, en efecto, no puede dejar fuera las dimensiones de cobertura y eficiencia, aunque no pueda limitarse a ellas y deba incluir además el nivel de aprendizaje, entre otras cosas.

Un concepto amplio de calidad debe incluir, en mi opinión, varias dimensiones que, en forma sintética, pueden expresarse diciendo que un sistema educativo de calidad es aquél que:

- Establece un currículo adecuado a las necesidades de la sociedad, incluyendo las de una mayor productividad económica, pero también otras necesidades básicas en una perspectiva de desarrollo integral, como las que tienen que ver con la democracia política, el respeto de los derechos humanos, el desarrollo de la ciencia, el cuidado del medio ambiente y la preservación y enriquecimiento de la diversidad cultural. Esta dimensión puede definirse con la expresión relevancia de los objetivos curriculares.
- Logra que la más alta proporción posible de destinatarios acceda a la escuela, que permanezca en ella hasta el final del trayecto previsto y que egrese alcanzando los objetivos de aprendizaje establecidos. Esta dimensión incluye, pues, la cobertura y la eficiencia terminal, así como el nivel de aprendizaje, y en términos sistémicos coincide con la eficacia interna del sistema.
- Consigue que los aprendizajes logrados por los alumnos sean asimilados por éstos en forma duradera y deriven en comportamientos sociales sustentados en los valores de libertad, equidad, solidaridad, tolerancia y respeto a las personas, que son fructíferos para la sociedad y para el propio individuo, quien podrá así alcanzar un desarrollo pleno en los diversos roles que habrá de desempeñar como trabajador, productor,

consumidor, padre de familia, elector, servidor público, lector y telespectador, entre otros; en pocas palabras, como ciudadano cabal. En términos sistémicos esta dimensión es la de la eficacia externa o impacto del sistema.

- Cuenta, para lograr lo anterior, con recursos humanos y materiales suficientes, y los aprovecha de la mejor manera posible, evitando despilfarros y derroches. Es fácil apreciar que esta es la dimensión eficiencia del sistema.
- Tiene en cuenta la desigual situación de alumnos y familias, de las comunidades en que viven y las escuelas mismas, y ofrece apoyos especiales a quienes lo requieren, para que los objetivos educativos sean alcanzados por el mayor número posible. Esta última dimensión es la de equidad.

Conceptualizada así, multidimensionalmente, la educación será de calidad si es relevante, si tiene eficacia interna y externa, si tiene un impacto positivo en el largo plazo, si es eficiente en el uso de los recursos y si busca la equidad.

Considero que es mejor este concepto amplio de calidad que uno restringido, porque creo que nadie creará que un sistema educativo es de calidad si atiende sólo a una fracción de la población en edad de acceder a cierto nivel, o si pierde en el camino a muchos de los alumnos que lo comienzan, aunque los que terminan lo hagan con altos niveles de aprendizaje.

Los otros componentes de una buena evaluación son claros: todas las áreas del currículo, comenzando por supuesto con Lecto-escritura y Matemáticas, pero sin limitarse a ellas, sino contemplando, al menos en el mediano plazo, las ciencias naturales y las del hombre, la educación física y artística, etc.

Deberán incluirse los aspectos cognitivos pero también los actitudinales y valorales, ya que la educación no sólo pretende desarrollar los primeros, sino también, y con creciente énfasis, los segundos, cuya importancia para la vida en una sociedad democrática y culta se reconoce cada vez más. Obviamente, no sólo deberán valorarse niveles de aprendizaje elementales, como memorización, sino también los más elevados, de razonamiento propio, juicio crítico y creatividad.

Una buena evaluación no podrá limitarse a los insumos del sistema, si bien no puede olvidarlos, ya que sin ellos no es posible medir la eficiencia del sistema, pero habrá de considerar también los productos, para valorar la eficacia, y los procesos, como parte de una aproximación explicativa que indague en lo relativo a las causas de la situación descrita, con lo que se dará sustento sólido al diseño y la implementación de acciones de mejoramiento.

Con un enfoque integral como el que se propone, la evaluación tomará en cuenta no sólo a los alumnos, sino también a los maestros, los directores y supervisores y las autoridades educativas, así como los diferentes niveles de organización de un sistema complejo como

el educativo, en cada uno de los cuales hay una problemática específica, y oportunidades particulares de intervención: el aula, la escuela singular, la zona o distrito escolar, el municipio, la región, provincia o entidad federativa, y el país.

Una buena evaluación, además, deberá caracterizarse por las cualidades técnicas de cualquier medición buena, que se resumen en la validez y la confiabilidad, en sus diversas facetas: que se mida realmente lo que se pretende medir, y que se haga de manera que los resultados sean comparables en el espacio y en el tiempo.

Subrayo que la evaluación no puede reducirse a la aplicación de pruebas de aprendizaje. Estas son, sin duda, necesarias para la medición de los resultados de la enseñanza, pero la valoración de otros aspectos implica la construcción de indicadores con otro tipo de datos, tomados de otras fuentes, como los sistemas de información estadística o estudios ad hoc.

Es claro que esta es una visión ideal, que en sentido estricto es imposible llevar a la práctica, pues exigiría una cantidad de trabajo y un volumen de recursos enorme para recolectar la información necesaria.

Pero si no se tiene este punto de referencia es fácil caer en el error opuesto, que en la realidad se ha dado en muchos lugares: el de reducir el sistema de evaluación a los elementos más fácilmente disponibles, con lo que sólo se dispone de datos gruesos y burdos sobre el número de alumnos inscritos en el sistema al inicio y al fin de un ciclo, el de maestros y escuelas, el presupuesto asignado a las escuelas y poco más.

Por ello creo que es importante manejar una visión ideal de lo que debería ser la evaluación como referente para el desarrollo de un sistema concreto; con tal punto de referencia, y con ayuda de las metodologías de investigación y evaluación educativa desarrolladas en las últimas décadas, es factible diseñar sistemas mucho más finos que los habituales, con requerimientos de recursos al alcance de los sistemas educativos reales.

No está de más precisar que diversos tipos de evaluación son competencia de diversas instancias. En particular conviene distinguir la evaluación de individuos de la de los sistemas como tales. La evaluación de los alumnos en lo individual, por ejemplo, es y sin duda seguirá correspondiendo a los maestros y las escuelas en que están los alumnos; la evaluación de los maestros en lo individual corresponde y corresponderá seguramente a los directores y supervisores.

La evaluación a la que me refiero en este escrito es, precisamente, la del sistema educativo en cuanto tal. Se trata, pues, de una evaluación de nivel macro, que no sustituye a las evaluaciones individuales, sino que las complementa, iluminando un ángulo diferente de la realidad educativa.

Tras lo anterior, las respuestas a las siguientes preguntas serán más breves.

¿Qué estamos evaluando?

Resumo mi respuesta, en forma algo simplista y cruda, diciendo que evaluamos algunas de las dimensiones de la calidad educativa, pero dejando fuera varias de las más importantes, que lo hacemos de manera imprecisa, y que no llegamos a las causas.

Hasta hace muy pocos años, la situación de la mayor parte de los sistemas de evaluación educativa de los países de la región era justamente la que se ha mencionado antes: solamente se contaba con datos sobre alumnos, maestros, escuelas y presupuestos, con lo que se podían estimar indicadores de cobertura, eficiencia terminal y costo alumno en forma gruesa, ya que por lo general no se controlaba la calidad de los procesos de recolección de información, no se manejaban datos por edad, con lo que sólo podían construirse tasas brutas y no netas y se manejaban sólo cohortes aparentes.

Los esfuerzos que han hecho muchos de nuestros países en la última década han incluido, por una parte, mejoras en los sistemas de información y, por otra, el desarrollo de sistemas de pruebas nacionales para valorar los aprendizajes de los alumnos. Estos esfuerzos son importantes, sin duda, pero aún están por consolidarse y adolecen de fallas claras, explicables por la falta de una tradición en el campo y la consecuente escasez de recursos humanos especializados.

Pecando tal vez de simplismo, creo que en la región atendemos la cobertura y la eficiencia terminal en forma imperfecta y el nivel de aprendizaje de los alumnos en áreas curriculares y grados clave, pero por lo general limitándose a niveles cognitivos bajos. La contraparte obvia, lo que no estamos evaluando, incluye áreas curriculares adicionales, niveles cognitivos superiores, aspectos actitudinales y valorales, procesos tanto pedagógicos como de gestión y lo relativo a equidad.

En las escuelas se evalúan muchas de estas cosas, desde luego, pero no parece haber procedimientos para valorar tales dimensiones o aspectos de la calidad de manera permanente y sistemática en el nivel nacional. Debe añadirse que muchas veces las evaluaciones que hacemos no tienen, en un grado suficiente, las características técnicas mencionadas de validez y confiabilidad.

Por lo que se refiere a validez, la ausencia de controles de las variables relevantes del contexto puede hacer que midamos en realidad pobreza, cuando creemos medir calidad de las escuelas. En este sentido en la región es muy raro todavía el uso de las poderosas técnicas actuales, como los modelos jerárquicos lineales, que permiten identificar la influencia de los factores del entorno distinguiéndolos, por ejemplo, de los de la escuela, el aula y el propio alumno.

En cuanto a confiabilidad, es también raro el uso de metodologías psicométricas avanzadas, como los modelos de la teoría de la respuesta al reactivo, que permiten contar con escalas más precisas, comparables en el tiempo, que cuidan la unidimensionalidad y evitan el sesgo de manera más adecuada que las técnicas de la teoría clásica de las pruebas. Debe añadirse que, hasta hace poco tiempo, no era raro que se manejaran pruebas que

no empleaban en forma rigurosa siquiera los principios de la teoría clásica. Las técnicas de valor agregado son también muy poco conocidas.

¿Qué de lo evaluado resulta significativo?

En principio los mecanismos de evaluación existentes son todos significativos y, aun en ausencia de los que se sugiere añadir, contienen potencialmente elementos útiles para el mejoramiento de los sistemas educativos. Sin embargo, es frecuente que aún esos elementos se desaprovechen y sólo sirvan para acumular datos en anuarios estadísticos destinados a empolvase en los estantes, sin que los responsables de tomar decisiones, y mucho menos los maestros, los conozcan y utilicen para retroalimentar su quehacer.

Esto quiere decir que, además de mejorar los actuales mecanismos de evaluación y de enriquecerlos con nuevos elementos, es necesario un importante esfuerzo de difusión de resultados y de capacitación de los usuarios para que estén en condiciones de entender y aprovechar los resultados de las evaluaciones.

¿Por qué las reformas realizadas en los 90 no se ven traducidas en los resultados?

Esta pregunta parte de un supuesto que no me parece incuestionable: que las reformas educativas implementadas en América Latina en los '90 no han dado resultados.

Los niveles educativos de nuestros países son el resultado de un enorme conjunto de factores, incluyendo unos tan complejos como el explosivo incremento demográfico que hizo del subcontinente la región de mayor crecimiento poblacional en toda la historia hasta mediados del siglo XX, la pobreza que aflige a gran parte de la población y la diversidad étnica y lingüística. Otros factores igualmente complejos incluyen el carácter autoritario de muchos regímenes políticos, las estructuras sindicales y burocráticas, y la desigualdad ancestral de nuestras sociedades.

Los cambios educativos, por otra parte, son complejos también y, en el mejor de los casos, llevan mucho tiempo; no pueden ser el resultado rápido de esfuerzos simples.

Por ello, aun si en verdad las reformas de los años 90 no han dado resultados, podrían encontrarse diversas explicaciones.

Las políticas educativas de la región han adolecido, en muchos casos, de serias deficiencias en diversos aspectos: en el diagnóstico en que se sustentan, muchas veces incompleto e impreciso; en el diseño de los programas, muchas veces simplista, apostando demasiado a pocos elementos de los que se espera más de lo que pueden dar; y, tal vez, sobre todo, en la implementación.

No es fácil, en efecto, que las reformas lleguen hasta el nivel más importante, el de la escuela y el aula. Para ello se requiere de tiempo y, además, de estrategias muy finas que aseguren la participación de los maestros. Desgraciadamente, es más frecuente encontrar estrategias simplistas e implementaciones apresuradas que procesos cuidadosos y respetuosos de los actores de base y de la complejidad de la cuestión.

No debería sorprender, por consiguiente, que los resultados obtenidos disten mucho de los esperados.

Pero el punto que quiero destacar es que no tenemos evidencia suficiente para afirmar de manera contundente que las reformas de los '90 no han tenido resultados. Hay, sin duda, indicios importantes en tal sentido, pero los sistemas de evaluación de que disponemos no sustentan afirmaciones tajantes.

En particular, creo que en México hay elementos para sostener que el nivel de aprendizaje ha mejorado ligeramente en los últimos años; que el impacto del gasto público en educación básica está siendo mayor para los tres deciles más pobres de la población, contra lo que ocurría antes; y que los programas compensatorios emprendidos a principios de los 90 han logrado reducir la deserción en medio rural y mejorar los niveles educativos en las zonas más desfavorecidas.

Por lo demás, la gran desigualdad que caracteriza a nuestras sociedades hace que necesitemos evaluaciones más precisas, que tengan en cuenta el peso de los factores del entorno, tanto en razón del principio de equidad, como para valorar en su justa medida los avances que puedan darse en contextos diferentes, en los que la mejora es más o menos difícil.

Conclusión

La última pregunta es, a mi juicio, la más importante de las cuatro que he tratado de responder, tanto por lo que expresamente plantea, como por lo que sugiere de manera implícita.

Ampliando lo dicho hasta ahora trataré de aportar algunas ideas al respecto.

La experiencia de las reformas educativas implementadas en los países de mayor desarrollo económico ha dejado lecciones importantes, en la dirección ya apuntada: no deben esperarse cambios espectaculares; las mejoras educativas se consiguen en lapsos de tiempo largos, y mediante esfuerzos sistemáticos y complejos, no gracias a recetas mágicas simples o panaceas.

Hoy sabemos también que no debe exagerarse la importancia de los recursos materiales; estos son importantes, sin duda, sobre todo hasta ciertos umbrales mínimos, y tal vez más en cuanto a la regularidad o continuidad de su existencia que en cuanto a su monto absoluto. Pero parece claro que tanto y más importan los procesos a través de los que se utilizan los recursos: procesos pedagógicos especialmente en el interior del aula, procesos más amplios de gestión en el nivel nacional y regional y, sobre todo, en cada escuela, donde el trabajo del equipo docente, el liderazgo del director y la participación de los padres de familia han mostrado ser fundamentales.

Para poder responder con bases sólidas a la pregunta sobre el éxito o el fracaso de las reformas educativas necesitamos, pues, en mi opinión, contar con sistemas de evaluación bastante más sólidos, aunque no mucho más costosos que los existentes, que nos den la información clave, de que hoy carecemos, sobre las diferentes dimensiones de la calidad educativa de nuestros países en forma continuada y permanente.

Es posible que nos llevemos algunas sorpresas agradables, en el sentido de que los esfuerzos realizados hasta ahora han dado algunos resultados, no espectaculares, pero sí razonables, en una perspectiva histórica y comparativa en el plano internacional.

También es posible que se confirme la idea de que no ha habido resultados positivos. En este caso, si tenemos buenos sistemas de evaluación tendremos también elementos para saber por qué ha ocurrido tal cosa, y para diseñar programas de mejora más adecuados.

Segunda Parte.

EL OPERATIVO NACIONAL DE EVALUACIÓN ONE 2009

MARCO DE REFERENCIA

a) Introducción

Contar con estrategias para mejorar la calidad de la educación implica, entre otros requisitos, disponer de un sistema de evaluación que haga explícito su marco de referencia y que permita conocer qué están aprendiendo los estudiantes.

Todo tipo de evaluación -y en especial las educativas-, exige modelos teóricos precisos que la justifiquen y la respalden. De este modo, la evaluación resulta sostenida por nexos conceptuales sólidamente relacionados, que permiten entrelazar los distintos pasos del proceso involucrado, desde la delimitación del objetivo de la evaluación a la posterior interpretación de los hallazgos. La selección de la metodología a llevar a cabo -que incluye de modo fundamental el tipo de diseño de los instrumentos, el proceso de implementación y el análisis posterior de los datos-, debe ser consecuente con este marco general.

Existen diferentes enfoques para la elaboración de pruebas que tienen como objetivo la evaluación de aprendizajes y existen también distintas teorías que sustentan su medición, así como diversos diseños entre los cuales optar para construir las pruebas de evaluación. Los siguientes apartados tienen como objetivo presentar, sintéticamente, el enfoque, la teoría de la medición y el diseño, adoptados como marco de referencia del Operativo Nacional de Evaluación (ONE) 2009.

b) Sobre el enfoque de la medición

Los enfoques vigentes, para abordar operativos de evaluación educativa, a nivel nacional e internacional, son dos, y se diferencian en cuanto al eje de referencia que cada uno adopta tanto para la elaboración de las pruebas como para la interpretación posterior de los resultados obtenidos. Desde este modo, este tipo de mediciones se dividen entre aquellas **referidas al grupo normativo** y las **referidas al criterio**. No existe una connotación de valor alguna asociada a cada uno de estos enfoques. Optar por uno u otro viene dada por el propósito de la evaluación.

Es digno de aclarar que, frecuentemente, se caracterizan a las pruebas como referidas al grupo normativo o al criterio, sin embargo, esta terminología es inapropiada. No es la prueba, sino el marco de referencia para interpretar el desempeño en la prueba, lo que se refiere al criterio o a la norma; de hecho, ambos tipos de interpretación pueden utilizarse en una prueba determinada. Luego de subrayar estos conceptos, veamos, concreta-

mente, las características fundamentales de cada uno de estos enfoques.

Cuando un estudiante es evaluado en el dominio de algún contenido y capacidad, a partir de mediciones referidas al grupo normativo, el puntaje obtenido por el alumno se interpreta por comparación con la norma de su grupo. De esta manera, las mediciones normativas no brindan información respecto a cuánto sabe y cuánto no sabe, sino que proveen datos acerca de su posición relativa con respecto a un grupo.

Este enfoque supone, como condición previa, que la prueba sea aplicada a un grupo normativo -representativo, un grupo de individuos con características similares a los futuros destinatarios de dicha prueba-, lo cual permitirá luego una lectura del rendimiento de cualquier otro estudiante. En otras palabras, la interpretación del o los puntajes obtenidos por un alumno en particular, es realizada en función de las puntuaciones obtenidas por un grupo de individuos, las cuales se constituyen como referencia para la interpretación de los puntajes alcanzados por otros alumnos.

Para analizar las consecuencias prácticas de la aplicación de este enfoque en educación, y a modo de introducción a otro, alternativo, tomemos un ejemplo sencillo. En principio, imaginemos que a un grupo, seleccionado como representativo, se le administra una prueba, y éste realiza de modo correcto, en promedio, el 30% de dicha prueba. Luego, un alumno en particular, contesta correctamente el 50% de la misma, que, por comparación con el grupo normativo, puede traducirse en una calificación de 8 puntos. Dentro de este contexto, supongamos también, que el docente valore como "insatisfactorio" que el alumno haya contestado sólo el 50% de la prueba, ya que, a partir del proceso de aprendizaje en cuestión, espera un dominio mayor de los contenidos y capacidades evaluados. Surge entonces una discrepancia entre la calificación 8 (ocho), que señala un rendimiento muy bueno, y el criterio pedagógico, que valora los conocimientos como insatisfactorios. Dicha discrepancia alude al núcleo de las diferencias entre el enfoque referido a la norma -comparación- y el referido al criterio -pedagógico-.

Si bien la evaluación referida a normas resulta muy útil en la medición de numerosas variables, es obvio que no es la más adecuada para valorar los aprendizajes. En el contexto educativo, y especialmente en la evaluación de los aprendizajes, es conveniente la medición referida a criterios, orientada a partir de la definición de las capacidades y contenidos en juego. En las evaluaciones educativas, los contenidos comunes de los currículos oficiales -los cuales identifican los dominios conceptuales

y los procesos cognitivos involucrados en los distintos niveles educativos-, constituyen la referencia de los criterios adoptados.

Resumiendo, tanto la elaboración de instrumentos como la interpretación de los resultados referida a criterios, es, sin duda, la más adecuada en el ámbito educativo, ya que es el enfoque más congruente con los propósitos de las evaluaciones que se realizan en esta área.

Consecuentemente, a partir de 2005, se ha seleccionado este enfoque de la medición como marco de las evaluaciones llevadas a cabo a través de los Operativos Nacionales de Evaluación (ONE), ya que permite determinar si los alumnos han logrado un nivel predeterminado, a partir de un criterio prefijado, en una disciplina dada.

Los instrumentos derivados de este tipo de enfoque requieren descripciones detalladas de lo que miden, con el fin de brindar información precisa sobre los logros respecto de un estándar. En la escuela y más allá de ella, la información provista por instrumentos elaborados con este enfoque resulta útil, en principio, a maestros que desean mejorar sus prácticas instrucción, a directores e inspectores que necesiten apoyar a los profesores y a los diseñadores tanto de políticas para mejorar las escuelas, como a los diseñadores de currículum y de textos que requieran verificar la efectividad de ambos, currículum y textos.

Desde el punto de vista de las distintas fases de construcción de una prueba basada en este enfoque, la primera y fundamental, es aquella que tiene como objetivo explicitar el criterio lógico de selección de los ítems, teniendo en cuenta que el más adecuado será aquel que permita separar mejor a los examinados de acuerdo con sus niveles de desempeño en los distintos dominios. Es digno de subrayar que los criterios deben ser definidos a priori, y que estas definiciones son las que permitirán el cierre del proceso de evaluación, al permitir la interpretación de los resultados de los estudiantes examinados en términos de logros del dominio o la aptitud en cuestión.

c) Acerca de la teoría de la medición

Toda medición científica se halla fundamentada en una teoría de la medición que tiene como objetivos proveer fundamentos conceptuales que permitan establecer asociaciones entre diversos sistemas formales y los atributos que interesa medir; mejorar la utilidad y la precisión de las mediciones; analizar distintas propuestas para describir, categorizar y evaluar la calidad de las medidas y desarrollar nuevos métodos en la obtención de instrumentos de mayor calidad.

La necesidad de este sustento teórico es particularmente relevante en disciplinas como la psicología y la educación, que se interesan por medir aspectos que no son directamente observables, y que, por tanto, deben inferir atributos subyacentes a partir de los comportamientos observables de los examinados.

En la actualidad, en términos generales, se dispone de dos teorías que sustentan la construcción de pruebas psicométricas: la Teoría Clásica de los Tests (TCT) y la

Teoría de Respuesta al Ítem (TRI).

Un test siempre se propone establecer inferencias sobre los rasgos psicológicos (no observables) de los sujetos basándose en la información que manifiestan en las respuestas. Tanto la TRI como la TCT consideran que cada sujeto lleva asociado un parámetro individual. En la TRI se denomina aptitud, incluyendo cualquier rasgo psicológico y se simboliza por la letra griega (zeta), mientras que en la teoría clásica se denomina puntaje verdadero (V).

La TCT es una teoría útil para describir la influencia de los errores de medida en las puntuaciones observadas u obtenidas a través de instrumentos, y sus relaciones con las puntuaciones verdaderas. Se basa en el Modelo Lineal de Spearman, desarrollado a principios del siglo XX. Se trata del primer modelo que aborda el problema de la incertidumbre o error inherente a cualquiera de las medidas realizadas mediante la aplicación de un test. Aún cuando posteriormente se han desarrollado nuevas teorías -tales como la TRI-, la TCT continúa vigente.

La diferencia principal entre la TCT y los diversos modelos basados en la TRI, es que la relación entre el valor esperado y el rasgo o aptitud, en la primera es de tipo lineal, mientras que en los modelos de la TRI las relaciones pueden ser funciones de tipo exponencial, tales como los modelos de Poisson, de la ojiva normal, del error binomial, el modelo de Rasch o los modelos logísticos de 1, 2 o 3 parámetros. Existen numerosos criterios con los cuales comparar la TCT con la TRI, muchos de ellos, como los mencionados arriba, son muy técnicos. Numerosas publicaciones dan cuenta de ellos, en el presente trabajo se prefiere más bien hacer referencia a algunos aspectos conceptuales.

Sintéticamente, señalemos que la limitación más importante de los tests elaborados según la TCT es que no permiten separar las características del examinado de las características del test: cada uno de estos aspectos sólo puede ser interpretado en el contexto de la interacción entre unas y otras. En concreto, la aptitud de un examinado se define en términos de un test o prueba específica. Si el test es "difícil" el examinado aparecerá como de poca aptitud, si el test es "fácil" el examinado parecerá tener mucha aptitud. La dificultad de un ítem se define, según la TCT, como la proporción de examinados que contesta el ítem correctamente en un grupo determinado. A su vez, las características métricas de la prueba, tales como la confiabilidad y la validez, se definen también en términos de un grupo determinado de examinados a partir del cual se construye el baremo o las normas de interpretación de las puntuaciones. Esto implica que es muy difícil comparar los resultados de examinados que han sido examinados con distintas pruebas.

Para subsanar estas limitaciones, en la década de 1960, numerosos investigadores comenzaron a difundir y aplicar otros métodos de control experimental, el del ajuste estadístico. Este último requiere la parametrización explícita de la aptitud que interesa evaluar, así como de las propiedades de los ítems, según un modelo que relacione sus valores con los datos de las respuestas relevadas a través de la aplicación de la prueba. Si el modelo se sostiene y los parámetros de los ítems se co-

nocen, el modelo ajusta los datos según las propiedades de los ítems del test y por lo tanto puede ser usado para producir mediciones de la aptitud que están libres de las propiedades de los ítems de la prueba aplicada.

La puntuación de una prueba en el modelo clásico estima el nivel de un atributo (aptitud, dominio de una capacidad evaluada a través de determinado contenido) como la suma, más estrictamente, la sumatoria, de respuestas a ítems individuales, mientras que la TRI utiliza el patrón de respuesta. De este modo, la TRI constituye el marco adecuado para el análisis de los aspectos vitales de las pruebas educativas construidas con el enfoque referido al criterio. Permite evaluar, sobre el dominio de la escala de aptitud, la separación efectiva entre niveles.

La TRI posee ventajas considerables con relación al paradigma clásico, tales como las de generar medidas diferentes con ítems estrictamente comparables y no dependientes de las muestras específicas de investigación, así como alcanzar un verdadero nivel intervalar de medición. Este modelo posee también dificultades en su implementación, ya que requiere de muestras grandes –entre 200 y 500 sujetos- para calibrar los ítems, pero esta desventaja no resulta un problema en operativos de amplio alcance como los ONE.

Una aplicación sumamente útil, de la TRI es que posibilita la construcción de bancos de ítems, vale decir, un conjunto de ítem que miden una misma variable y cuyos parámetros están estimados en una misma escala. Estos ítems con sus parámetros se pueden almacenar y construir en los llamados bancos de ítems.

En resumen, la TRI intenta dar una fundamentación probabilística al problema de la medición de constructos inobservables. Su denominación proviene del hecho de considerar al ítem como unidad básica del test. Los modelos que utiliza son funciones matemáticas que relacionan las probabilidades de una respuesta particular a un ítem con la aptitud general del sujeto. Si bien su origen no es tan nuevo, dada la complejidad de los cálculos para su aplicación solo empezó a difundirse y utilizarse gracias a programas de computación específicos como BIGSTEP, LOGIST, BILOG, entre otros.

Como todo modelo matemático, la TRI incluye un conjunto de supuestos acerca de los datos en los cuales se aplica. En términos generales, consideran tres supuestos básicos:

Unidimensionalidad. Supone que cuando se diseña una prueba, una única habilidad es necesaria para explicar o dar cuenta del desempeño del examinado en la prueba. Si bien se reconoce que cuando un estudiante responde a un ítem en una prueba confluyen múltiples habilidades, los ítems deben diseñarse haciendo énfasis en una de ellas o en una combinación particular predefinida.

Independencia Local. Se espera que un estudiante evaluado responda a un ítem en particular sin que recurra a información de otros ítems para hacerlo correctamente. Es decir, la ejecución en un ítem no debe afectar sus respuestas en otro.

Curvas Características de Ítems (CCI). La CCI es una función matemática que relaciona la probabilidad de éxito en una pregunta con la habilidad, medida por el conjunto de ítems que la contienen. Todo ítem puede describirse por su CCI y es esta curva la unidad conceptual básica de la TRI.

Existen diferentes modelos de la TRI que se diferencian en la forma particular que adquiere la función de probabilidad, la cual constituye el número específico de parámetros a analizar. El modelo utilizado para el procesamiento y análisis de los datos del ONE 2009 es el Modelo de Rasch.

Por último, es digno de subrayar que las mediciones referidas al criterio (v. apartado 1), de antiguo uso, han recibido un nuevo impulso en el ámbito educativo dentro del marco teórico de la TRI. De este modo se señala la congruencia entre la elección del enfoque y la teoría adoptados en el ONE 2009.

d) Cuestiones relacionadas con el diseño

Los diseños que pueden aplicarse para llevar a cabo una evaluación de aprendizajes son muy variados. Capitalizando la experiencia adquirida en operativos anteriores e incluyendo novedades teórico-técnicas, a partir del ONE 2009, se ha optado por dividir en dos pruebas la evaluación de cada estudiante en un dominio específico, una de ellas constituida por reactivos de opción múltiple, de respuesta cerrada, y la otra por ítems que implican la construcción de respuesta, llamados también abiertos. Ambas se administran durante la misma jornada, con un breve intervalo de tiempo -receso, recreo-, entre ellas. El diseño seleccionado apunta a asegurar la mayor cobertura de las respuestas y de este modo evaluar una diversidad de contenidos y procesos cognitivos.

Concretamente, cada estudiante responde a dos cuadernillos (uno con ítems cerrados y otro con ítems abiertos), asignados en forma aleatoria, excepto los destinados a la evaluación del área Comprensión Lectora, en la cual se conserva un nexo prefijado entre un texto del Cuadernillo con ítems cerrados y los ítems abiertos.

El cuadernillo con los reactivos de opción múltiple cuenta con 30 ítems, mientras que los de construcción de respuesta están conformados por 2 ítems.

El proceso de construcción de los reactivos con respuesta cerrada ha sido llevado a cabo mediante la implementación de un diseño matricial.

Se trata de un tipo de diseño que se ha mostrado adecuado en evaluaciones educativas internacionales porque brinda soluciones a los problemas que intervienen cuando las variables a considerar son muy numerosas. Su instrumentación implica la elaboración de un conjunto amplio de ítems que cubran el dominio del currículo que se desea evaluar. Luego estos ítems se dividen en grupos de preguntas denominados bloques, de tal modo que cada uno evalúa de modo incompleto dicho dominio. De ahí, que, específicamente, el diseño presenta una estructura en cuadernillos con bloques incompletos balanceados, que en su conjunto permiten evaluar una amplia gama de dominios y procesos en los aprendizajes a evaluar.

De esta manera, las pruebas construidas con este diseño permiten evaluar un amplio rango de contenidos curriculares distribuyendo los contenidos de un dominio de contenido en distintas pruebas, aunque cada estudiante en particular resuelve cuestiones relacionadas con una parte del dominio (cada uno de formatos o modelos de la prueba, que se aplica a cada estudiante está conformado por dos bloques).

Cada bloque de ítems cerrados está conformado por 15 ítems. Los 8 bloques de ítems cerrados se distribuyen en forma de espiral a razón de dos por cuadernillo. De este modo resultan 8 Cuadernillos o Modelos de ítems cerrados y 4 Cuadernillos o Modelos de ítems abiertos.

Cada cuadernillo de administración, compuesto por ítems cerrados, consta, entonces, de dos de los bloques, que van alternando su posición a lo largo de distintos cuadernillos de tal forma que el que se aplicó en la primera posición aparezca en otro ocupando la segunda y viceversa, como se ejemplifica en la siguiente tabla.

Nombre del cuadernillo	Nombre del bloque en 1° posición	Nombre del bloque en 2° posición
PM3C1	M3_1	M3_2
PM3C2	M3_3	M3_4
PM3C3	M3_5	M3_6
PM3C4	M3_7	M3_8
PM3C5	M3_2	M3_3
PM3C6	M3_4	M3_5
PM3C7	M3_6	M3_7
PM3C8	M3_8	M3_1

En otras palabras, aunque cada alumno -a la hora de la aplicación del ONE 2009-, contesta sólo algunos reactivos, el diseño que enmarca el operativo de evaluación en su conjunto, provee información sobre la totalidad de las capacidades y contenidos de los dominios en cuestión.

Para la prueba ONE 2009, se han construido, para cada área, en cada nivel, la siguiente cantidad de ítems:

Tipo de ítem	Cantidad de ítems	Cantidad de bloques (15 ítems)	Cantidad de Cuadernillos
Abiertos	4	---	4
Cerrados	120	8	8

El procesamiento de la información aportada por la aplicación piloto de esta cantidad de ítems permitirá seleccionar las pruebas más confiables y válidas que serán administradas en el ONE 2009 definitivo.

“EVALUACIÓN. ENFOQUES METODOLÓGICOS”

Justificación para un cambio de enfoque en cuanto a los ONE: de pruebas referidas a normas a pruebas referidas a criterios. ¿Para qué?, ¿por qué? y ¿en qué consiste?

“Creo que llega el momento de empezar a llamar a las cosas por su nombre, nos hemos propuesto evaluar calidad y estamos midiendo, no evaluando. Evaluar exige tener un patrón con qué compararse y ese patrón no está del todo claro.”¹¹

Como señala la cita que inicia este trabajo, pensar en la evaluación de la calidad educativa “exige” tener un patrón con qué compararse, claro y específico. Así es como se ha adoptado la decisión política de redefinir la evaluación tal como se viene desarrollando, y se solicitó a los equipos técnicos-pedagógicos de la DINIECE, la elaboración de un nuevo diseño de pruebas para evaluar la calidad educativa.

A partir de una demanda bastante generalizada, proveniente de diferentes sectores, se torna relevante la búsqueda y explicitación de una definición de sentido y de valor sobre los procesos vinculados a la evaluación de la calidad educativa. Atendiendo a este propósito, se pensó en un cambio de enfoque para las futuras pruebas nacionales, capaz de generar y de brindar nuevos aportes y diferentes miradas para mejorar las formas de enseñanza y los aprendizajes de los alumnos.

Se debe reconocer el enorme esfuerzo realizado por la Argentina en todos estos años en materia de evaluación de los procesos de la calidad educativa y los logros alcanzados en este campo:

- La utilización de la Teoría Clásica de las Pruebas (TCP) llevado a cabo bajo el enfoque de Pruebas Referidas a la Norma (PRN).
- La producción de las Recomendaciones Metodológicas para el mejoramiento de las formas de enseñanza y los aprendizajes de los alumnos.
- El apoyo técnico-pedagógico brindado a los Ministerios de Educación Jurisdiccionales en el desarrollo y fortalecimiento de sus propios equipos técnicos.
- “Los sistemas de medición han convertido los resultados educativos en un motivo de debate público...” (Tedesco, 2003)

Por lo tanto, se propone seguir con el empleo de estos análisis para el procesamiento y la producción de resultados, a fin de construir la necesaria “continuidad” en todo proceso educativo y el “enriquecimiento” hacia la nueva propuesta.

Pero para resignificar, en términos del sentido y del valor, los procesos de evaluación de la calidad educativa se hace necesario “revisar” los criterios políticos y técnicos-pedagógicos asumidos a lo largo de estos años, y desde allí, a partir del disenso y el consenso entre los diferentes actores, construir nuevos y/o complementarios criterios de evaluación enmarcados en un contexto diferente y regidos por la crítica, la reflexión y la toma de conciencia en esta compleja problemática, a fin de lograr una “apropiación” por parte de los actores educativos de los objetivos propuestos.

Durante muchos años, en la DINIECE se utilizó la Teoría Clásica de las Pruebas (TCP) para diseñar instrumentos de evaluación. Los resultados se presentaron en términos de porcentaje de respuestas correctas. El porcentaje medio de aciertos expresa cuál es el tanto por ciento medio de aciertos de los alumnos en los ítem de las distintas pruebas, pero no lo que los alumnos saben o no saben. No es válido identificar el éxito o fracaso en función de un valor medio de aciertos, y no existe valor alguno que, a priori, pueda considerarse como rendimiento satisfactorio o insatisfactorio. Con todo, la presentación de los resultados en términos de porcentajes facilita su interpretación por el lector no especializado.

Hoy, el nuevo diseño de pruebas para evaluar la calidad educativa se inicia bajo dos ejes fundamentales: el enfoque de Pruebas Referidas al Criterio (PRC) y la utilización de la Teoría Respuesta al Ítem (TRI) como modelo matemático para el procesamiento y la producción de resultados. Los resultados se presentan de acuerdo con escalas de rendimiento para cada área evaluada. La escala, por ejemplo de 0 a 500, representa el resultado que habría obtenido un alumno que hubiera contestado a una teórica prueba de 500 ítems. Se pueden establecer puntos de corte en la escala y se definen las tareas que son capaces de realizar aquellos alumnos que alcanzan los respectivos niveles. De esta manera, se introduce un elemento criterial en la evaluación.

¹¹ Lafuente, Marta. En “Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa”. IIFE - UNESCO. Argentina 2003.

En el marco de este enfoque, la DINIECE se propone brindar información útil y específica sobre el logro de los objetivos fundamentales del currículo por los estudiantes que terminan un nivel o ciclo.

Con esta redefinición de la evaluación se persigue ampliar las posibilidades que ofrece la evaluación educativa, a través de pruebas de aplicación a grandes poblaciones. Se busca la posibilidad de tener estadísticos de ítem y de prueba que sean independientes de las poblaciones en las cuales se apliquen, de forma tal que se pueda realizar el seguimiento necesario de los resultados y tener una idea del efecto de ciertas políticas o programas.

La Teoría del Rasgo Latente, también denominada Teoría de Respuesta al Ítem, tuvo su origen hace varias décadas pero se desarrolla principalmente en los años ochenta, cuando la computadora irrumpe con fuerza en Estados Unidos, pues los cálculos exigidos para su aplicación resultan casi imposibles de poder ser llevados a cabo sin disponer de una plataforma informática para el tratamiento de los datos.

Esta teoría también tiene su origen en momentos en que la psicometría norteamericana reconoce las graves limitaciones que tiene la Teoría Clásica de las Pruebas. La principal de estas limitaciones tiene que ver con el hecho de que los resultados obtenidos con la aplicación de un ítem, de una prueba, están limitados por las circunstancias de aplicación, es decir, hacen referencia exclusivamente a las situaciones en que se contestan los ítems. Todos los estadísticos obtenidos en una aplicación concreta de una prueba son relativos a las características de la prueba y de los sujetos evaluados.

Como consecuencia de lo que acaba de ser señalado, esos estadísticos tienen un valor relativo, en cuanto a las muestras de ítem y los sujetos evaluados (Mariano Álvaro Page, 1993).

La Teoría de Rasgo Latente o Teoría de Respuesta al Ítem (TRI) parte del supuesto de que las puntuaciones obtenidas en un ítem por un sujeto –y por consiguiente en una prueba– dependen directamente del grado o nivel en que ese sujeto posee el rasgo medido. Este rasgo es “inobservable”, por lo que se lo denomina “rasgo latente” y es un constructo que se utiliza para explicar el hecho. Hay una relación directa entre el rendimiento observable del sujeto, las puntuaciones obtenidas en la prueba, y el nivel de posesión del rasgo no observable o latente. La relación entre las puntuaciones observadas y el nivel de posesión del rasgo latente se describen mediante una función matemática.

La teoría pone en relación el nivel de habilidad y la probabilidad de responder correctamente un ítem, por lo que puede definirse como la relación funcional entre el nivel de habilidad y la probabilidad de responder correctamente un ítem (Mariano Álvaro Page, 1993).

En términos generales, la TRI considera tres supuestos básicos (Hambleton y Swaminathan, 1985):

- **Unidimensionalidad:** en la TRI se asume que cuando se diseña una prueba ésta deberá medir, preferiblemente, una dimensión, una habilidad (el término habilidad, en su concepción psi-

cométrica, se refiere “al objeto de medición”). Se reconoce que cuando una persona responde a una pregunta en una prueba, entran en juego múltiples habilidades, pero las preguntas deben diseñarse haciendo énfasis en una de ellas o en una combinación de ellas en particular.

- **Independencia local:** se espera que un estudiante responda a una pregunta en particular sin que recurra a información de otros ítems para hacerlo correctamente. Es decir, la respuesta de un estudiante en una pregunta no debe afectar sus respuestas en otra. Es práctica generalizada elaborar pruebas en donde se diseñan ítem en relación con un contexto, del que dependen las respuestas del alumno; aquí también se aplica la independencia local entre los ítems y no entre ellos y el contexto. Es decir, el supuesto de independencia local implica que la respuesta de un sujeto a un ítem es independiente, estadísticamente, de las respuestas a los otros, pero no que los ítem de la prueba no estén relacionados.
- **Curva característica del ítem:** viene determinada por la función matemática entre el nivel de habilidad o rasgo latente medido y la probabilidad de responder correctamente al ítem.

A continuación se presentan algunos conceptos correspondientes al enfoque de las Pruebas Referidas al Criterio (PRC), a fin de fundamentar el nuevo diseño de pruebas para evaluar la calidad educativa.

Las Pruebas Referidas al Criterio (PRC), representan procedimientos para evaluar el rendimiento y/o conducta de los sujetos con relación a dominios de contenidos bien definidos, en vez de por referencia a la conducta de otros sujetos, como en las Pruebas Referidas a las Normas (PRN) (Martínez Arias, Rosario; 1995).

Cuando se quiere comparar el logro de ciertos grupos de estudiantes con los de otros, se puede trabajar dentro del enfoque de evaluación referida a normas, mientras que cuando se quiere conocer qué conocimientos o competencias específicas logran desarrollar los estudiantes se debe recurrir al enfoque de evaluación referida a criterios. Esta opción es la más conveniente desde el punto de vista pedagógico porque permite obtener información relevante acerca de los conocimientos, destrezas y habilidades específicas que un grupo de estudiantes logra dominar (Esquivel, Juan Manuel).

¿Cómo surgen las Pruebas Referidas al Criterio? (PRC)

La expresión “Pruebas Referidas al Criterio” apareció por primera vez en un artículo de Robert Glaser (1963) titulado “Instructional Technology and the Measurement of Learning Outcomes: Some Questions” y publicado en la revista *American Psychologist*, (Mariano Álvaro Page, 1993).

No obstante, a pesar de la aparente novedad del término en un momento en que la evaluación estaba dominada por el paradigma de las diferencias individuales, lo que representan las PRC y su finalidad ya había sido puesto de relieve muchos años antes. En 1913, Thorndike estableció una distinción similar a la de Glaser, que fue totalmente olvidada por el desarrollo de la psicometría clásica, especialmente en el intervalo comprendido entre las dos guerras mundiales.

En los años sesenta empiezan a comprobarse los resultados de la enseñanza programada y de otros programas de intervención y se vio que para su evaluación era preciso un diagnóstico previo de los sujetos antes de la intervención, en términos de lo que eran "capaces de hacer". Las pruebas tradicionales con su énfasis interpretativo en el rendimiento del grupo normativo, no servían para esta finalidad ya que no proporcionaban la información necesaria sobre el rendimiento del sujeto.

Con raras excepciones, como las que representan los trabajos de Hively (1968), Osburn (1968) y Popham y Husek (1969), puede decirse que no fue hasta el comienzo de los años setenta cuando comenzó a trabajarse de forma sistemática sobre este tipo de pruebas.

¿Qué son las Pruebas Referidas al Criterio (PRC) y cuáles son sus características?

Se debe señalar en primer lugar que no es fácil dar una definición de PRC, ya que no existe uniformidad ni siquiera en la denominación dada. La palabra criterio se refiere al dominio de contenidos o conductas.

La definición que hoy es más aceptada es la de Popham (1978): "un test referido al criterio se utiliza para evaluar el estatus absoluto del sujeto con respecto a algún dominio de conductas bien definido".

Hambleton y Rogers (1991) hacen una serie de precisiones a esta definición. En primer lugar, además de dominio de conductas, puede hablarse intercambiamente de objetivos, destrezas y competencias. En segundo lugar, el dominio debe estar bien definido, siendo variable la amplitud y los contenidos de este dominio, ya que éstos dependen de la finalidad de la prueba. En tercer lugar, cuando una PRC incluye más de un objetivo, los ítem que cubren cada uno de los objetivos suelen organizarse en subpruebas y el rendimiento de los sujetos es evaluado en cada uno de los objetivos. En cuarto lugar, es una práctica frecuente establecer estándares de rendimiento o puntos de corte, la definición de PRC no incluye explícitamente este requisito, ya que pueden darse interpretaciones meramente descriptivas del rendimiento de los sujetos, tales como que "Enrique ha respondido correctamente el 75% de las preguntas en la prueba de Geometría".

Otras precisiones:

- El número de objetivos medidos en PRC puede variar entre las diferentes pruebas.

- El número de ítem que miden cada objetivo es variable, pudiendo ser diferente el de distintos objetivos dentro de la misma prueba.
- No es necesario que el formato de los ítems sea de elección múltiple, aunque es éste el formato más frecuente.
- Una forma habitual para tomar decisiones del tipo "apto" "no apto" o "pasa" "no-pasa", es comparar la proporción de aciertos del estudiante con un criterio estándar de rendimiento, decidiendo que el sujeto "pasa" si su rendimiento es igual o mayor que el estándar y "no pasa" en caso contrario.

Los requisitos básicos para que una prueba pueda ser considerada PRC son los siguientes:

- La existencia de un conjunto de objetivos, contenido, capacidades, desempeños claramente definidos.
- Una proposición explícita de la finalidad de la prueba.

¿Cuáles son las diferencias entre las Pruebas Referidas a las Normas y las Pruebas Referidas al Criterio?

En apariencia hay pocas diferencias entre PRN y PRC ya que ambos tipos de pruebas suelen estar compuestas por ítem con formatos similares; requieren el mismo tipo de operaciones cognitivas de los sujetos, etc. No obstante, hay numerosas diferencias tanto en la construcción, como en la interpretación de las puntuaciones.

Las PRN y las PRC difieren en primer lugar en la finalidad de la evaluación. En las PRC el objetivo es determinar el estatus actual del rendimiento o conducta de un sujeto, o una clara descripción de sus características, normalmente con objeto de clasificarlo en una categoría determinada: pasa-falla; etc. Hambleton y Rogers (1991) encuentran una cierta analogía con las denominadas pruebas de diagnóstico.

En las PRN, la finalidad es describir al sujeto en el continuo de algún rasgo, expresando su posición relativa respecto al grupo de sujetos.

Así pues, en las pruebas referidas a un criterio, las inferencias que se hacen de las personas son siempre referidas a un determinado objetivo o fin, y se comparan las puntuaciones obtenidas con los objetivos que estaban previsto alcanzar, en un caso para saber si un sujeto ha alcanzado cierto nivel y, en otro, para saber si alcanza el nivel exigido para determinada tarea. En cambio, las pruebas referidas a la norma, la puntuación de cada sujeto se compara con las puntuaciones de los demás sujetos a los que se les ha aplicado la prueba, con las puntuaciones del grupo o muestra representativa de la población a la que pertenece.

En cuanto a la **construcción de la prueba**, ambos tipos establecen los elementos en términos de las conductas o rendimientos que el sujeto exhibe en algún momento determinado. En ambos casos se pone el acento en que las respuestas a valorar sean públicas, estandarizadas y fácilmente objetivables. En las PRN raramente se determina un dominio inicial de elementos, considerando únicamente la teoría del rasgo en la que se basa el autor de la prueba, para determinar la calidad de los ítems. En las PRC la determinación del dominio de contenido o especificaciones de contenido es central para la construcción de la prueba.

En cuanto a la **selección de los ítems**, las propiedades psicométricas consideradas para su inclusión en la prueba son diferentes. En las PRN el objetivo es maximizar las diferencias individuales, lo que lleva a la inclusión de un gran número de ítem de dificultad media para aumentar el poder discriminativo de la prueba, (para poner de relieve las diferencias individuales, debe maximizarse la varianza de la prueba, seleccionando ítem de dificultad media y alto poder discriminativo). En las PRC son los objetivos de la prueba los que marcan la norma para la inclusión de los ítems, así como el propósito y la finalidad del mismo. Por ejemplo, cuando se trata de pruebas de "apto" vs. "no apto" para la evaluación de los rendimientos en un programa particular, suelen incluirse elementos fáciles, respondidos por el 70 u 80% de los sujetos.

Por último, otra de las diferencias se encuentra en los **significados de la puntuación**, en una PRN las respuestas son consideradas como signos del rasgo latente que se pretende medir con la prueba. En las PRC la respuesta se considera un estimador de la conducta del dominio (Marines Arias, Rosario 1995).

¿Cómo se realiza el análisis de los ítems?

El aspecto más crítico en la construcción de un PRC después de la especificación del contenido y de los objetivos es la generación de ítem¹² que se conformen a las especificaciones del contenido y a los objetivos.

- La validez del contenido de los ítems: según el concepto de validez de contenido, la determinación de este tipo de validez requiere evaluar el grado en que la muestra de ítem de la prueba es representativa del contenido del dominio. Se han propuesto diversos procedimientos basados fundamentalmente en el juicio de expertos (Berk 1984).
- Puede clasificarse en dos tipos: a) congruencia ítem-objetivo o grado en que el ítem mide el objetivo que pretende medir, b) sesgos de contenido.
- Crocker y Algina (1986) recomiendan definir el dominio de interés, seleccionar un panel de expertos en el contenido, proporcionar un proce-

dimiento estructurado para emparejar los ítems con el dominio, recoger y resumir los datos.

- Análisis estadístico de los ítems: los datos obtenidos en el estudio piloto pueden utilizarse para evaluar la eficacia de los ítems en términos de si funcionan o no en la forma en que estaba previsto.
- Los pasos en este proceso son: seleccionar grupos criterios (se refiere a la selección de los estudiantes a los que se les aplicará la prueba); calcular los estadísticos del ítem (se refiere a el índice de dificultad, es la proporción o el porcentaje de personas que contestan correctamente el ítem y el índice de discriminación, mide los cambios en el rendimiento o las diferencias relacionadas con la instrucción); y realizar un análisis del posible sesgo de los ítem (serán eliminados todos aquellos ítem que usen un lenguaje sesgado referido al género, minoría étnica o subgrupo particular).

¿Cuál es la longitud de la prueba?

Hambleton, Hutten y Swaminthan (1976) en un estudio empírico en que comparan métodos de obtener los puntajes de dominio y su efecto en varios factores (entre ellos la longitud de la prueba) concluyen que un número de ítem igual a ocho da "suficiente base para evaluar el dominio del estudiante o para tomar decisiones de instrucción para los datos de pruebas con referencia a criterios".

Por su parte Popham (1978) afirma lo siguiente: "Para simplificar un poco, para muchas situaciones educativas en las que se emplearán pruebas con referencia a criterios, la prueba debe consistir de 10 a 20 ítem por dominio conductual".

Por otra parte, Sheehan y Davis (1979) recomiendan menos ítem por objetivo, ya que desarrollaron una batería de pruebas con referencia a criterios de matemática, en la que emplearon cuatro ítem por objetivo.

También, en las pruebas desarrolladas por Esquivel (1985) en matemática y ciencias se empleó tres ítems por objetivos. En general se determina que cuando las decisiones son formativas y para muestras de individuos el número de ítem varía entre tres y cinco por objetivo, mientras que decisiones sumativas e individuales requieren entre ocho y diez ítem por objetivo.

12 Para la construcción de ítem ver "Criterios para la elaboración, selección, supervisión, envío de ítem y texto" de la DINIECE.

¿Cómo se establecen los estándares o puntos de corte?

Finalmente, una de las principales funciones de las pruebas es proporcionar datos para la toma de decisiones, especialmente cuando se utilizan TRC. En estas situaciones suele ser necesario establecer un punto de corte para decidir acerca del rendimiento del sujeto en la prueba. La práctica de obtener los puntos de corte para determinar las capacidades mínimas se conoce en la evaluación educativa con el nombre de fijación de estándares. El estándar es un punto en la escala de las puntuaciones de la prueba que sirve para clasificar a los sujetos en dos categorías que reflejan diferentes niveles de capacidad, con relación a un dominio. Es frecuente que estas categorías se expresen en términos de "apto" o "no apto", "pasa" o "falla". Shepard (1984) considera que es el talón de Aquiles de los TRC.

El tema es tan importante que fue origen de un gran debate durante los años setenta, de tal forma que la revista "Journal of Educational Measurement" dedicó un número monográfico a esta cuestión en 1978. Glass realizó una revisión exhaustiva sobre los estándares y puntos de corte de la que concluye que todos los procedimientos son arbitrarios o que están basados en premisas arbitrarias (Martínez Arias, Rosario, 1995).

Clasificación de procedimientos

Ante la multiplicidad de métodos, fue necesario establecer algunos esquemas de clasificación. Meskauskas (1976) propuso dividirlos en "modelos de estado" y "modelos continuos", siendo los modelos de estado los que asumen que la capacidad es binaria, ya que el sujeto domina el tema o no lo domina; los modelos continuos asumen que el rasgo o constructo evaluado es continuo y que el papel de los métodos consiste en determinar el nivel del rasgo que determina un rendimiento competente.

Se han propuesto muchos modelos de estado, pero apenas se han usado en las aplicaciones prácticas, dominando este ámbito los modelos continuos. Hambleton y Eignor (1980) Hambleton y Rogers (1990) hacen una clasificación de éstos en tres tipos denominados "modelos empíricos, modelos de juicios y modelos combinados". Jaeger (1989) considera que en todos está implícito el juicio y prefiere clasificarlos en "modelos centrados en la prueba" y "modelos centrados en el rendimiento de los sujetos".

Descripción de los procedimientos

1) Basados en juicios sobre el contenido de la prueba: son un conjunto de procedimientos basados en juicios sobre los elementos de la prueba. Aunque se han desarrollado muchos, los que mayor trascendencia han tenido son los de Nedelsky (1954), Angoff (1971), Ebel (1972) y Jaeger (1978).

a) Método de Nedelsky: definió el estándar absoluto en términos de "conocimientos y destrezas que un sujeto posee para pasar de curso". Los define a partir de los elementos de la prueba de elección múltiple, en términos de la verosimilitud de que los sujetos sean capaces de eliminar respuestas incorrectas.

b) Método de Angoff: el procedimiento requiere que los jueces de la muestra de expertos examinen cada elemento de la prueba y estimen la probabilidad de que una persona mínimamente competente responda a cada ítem correctamente. Angoff sugiere que esta probabilidad es más fácil de estimar si los jueces piensan en términos de un amplio número de individuos mínimamente competentes y estiman qué proporción de este grupo respondería a cada elemento correspondiente. La suma de proporciones representaría la puntuación mínima aceptable. El método Angoff hasta el momento es muy utilizado, investigado y recomendado.

c) Método de Ebel: propuso un sistema similar al Angoff basado en los juicios acerca de los ítems, pero teniendo en cuenta además la relevancia de contenido del ítem y su nivel de dificultad. Para la valoración de cada ítem sugiere usar cuatro niveles de relevancia –esencial, importante, aceptable y cuestionable– y tres niveles de dificultad –fácil, medio y difícil–

d) Método de Jaeger: intenta hacer la tarea de los jueces más simple y concreta. Cada juez debe responder con Sí o No a la siguiente pregunta, para cada ítem "¿será capaz de responder correctamente a este ítem un sujeto que reciba la calificación de apto en un examen de...?" y "si no responde al ítem correctamente, ¿se le negaría el diploma? Sí... No...".

A pesar de los avances técnicos, todavía existen dificultades importantes en el establecimiento de estándares, Angoff expresa "hemos observado que los diversos procedimientos de juicio para establecer estándares, no solamente plantean problemas de acuerdo entre sí, sino también de obtener los mismos resultados en la aplicación repetida.

2) Métodos basados en el rendimiento de grupos de validación: muchos autores sugieren que el criterio de validez para evaluar un estándar debe determinarse juzgando los rendimientos de los sujetos examinados.

- a) Método del grupo límite.
- b) Método de los grupos de contraste.
- c) Otros métodos empíricos.

3) Métodos basados en los errores en las decisiones: Van der Linden (1980) "no es una técnica para establecer estándares, sino una técnica para minimizar las consecuencias de los errores de medida y de muestreo". Son modelos continuos que asumen una variable latente continua con un estándar predeterminado.

ANEXOS

Anexo 1

¿Cómo se construye una Prueba Referida al Criterio?

Hambleton y Rogers (1991) proponen 11 dimensiones en la construcción de una PRC. Dependiendo de la finalidad de la prueba se tendrá en cuenta a cada una de las dimensiones.¹³

1 / Planteamientos iniciales.

- Especificar el propósito de la prueba.
- Especificar los objetivos, contenidos, capacidades, desempeños que la prueba pretende evaluar.
- Especificar los grupos a los que se aplicará la prueba: características sociodemográficas, lingüísticas, etc.
- Decidir el formato de los ítems.
- Determinar el tiempo y los recursos económicos que se pueden dedicar a la construcción de la prueba.
- Identificar y seleccionar jueces expertos en el contenido de la evaluación.
- Especificar la longitud de la prueba y el tiempo aproximado de cumplimentación.

2 / Revisión de los objetivos.

- Revisar las descripciones de los objetivos para determinar si son o no aceptables.
- Seleccionar el conjunto definitivo de objetivos que evaluará la prueba.
- Preparar especificaciones de contenidos para los ítems de cada objetivo y revisarlos en cuanto a su claridad, seguridad.

3 / Redactar los ítems.

- Construir un número de ítem considerablemente mayor que el definitivo para el estudio piloto.
- Introducir los ítems en un banco de ítem.
- Edición de los ítems.

4- / Evaluación de la validez de contenido.

- Identificar un grupo de expertos en el contenido de la prueba y especialistas en evaluación.
- Examinar el grado de emparejamiento entre ítem y objetivos, representatividad del contenido, etc.
- Revisar la adecuación técnica de los ítems.
- Revisar de nuevo los ítems.
- Escribir ítems adicionales si es necesario.

5 / Aplicación de la prueba.

- Preparar una prueba piloto.
- Administrar la prueba a muestras adecuadas.
- Realizar el análisis de ítem y estudio del sesgo¹⁴.

6 / Revisar la prueba.

Usar los resultados de 5- c. para la revisión.

7 / Preparar la forma definitiva de la prueba.

- Determinar la longitud de la prueba, número de formas necesarias y número de ítem por objetivo.
- Seleccionar ítem del conjunto analizado.
- Preparar instrucciones, ejemplos, cuadernillos, hojas de respuestas, claves de corrección, etc.
- Especificar posibles modificaciones a las instrucciones, tiempo para alumnos con necesidades especiales, etc.

8 / Establecimiento de un estándar o punto de corte.

- Determinar si es necesario establecer un estándar de rendimiento, si no lo es, pasar a 9.
- Establecer el o los estándares, si habrá más de dos categorías.
- Especificar modificaciones a los estándares para posibles casos especiales.
- Especificar interpretaciones alternativas de las puntuaciones de la prueba para los casos de c.

9 / Aplicación en forma piloto de la prueba definitiva.

- Diseñar la aplicación de la prueba para recoger datos acerca de la fiabilidad y validez.
- Aplicar la prueba a muestras representativas de alumnos.
- Evaluación del procedimiento de administración, ítem, fiabilidad y validez.
- Hacer las revisiones finales a partir de los datos anteriores.

10 / Preparación del manual de instrucciones.

- Manual para los aplicadores de la prueba.
- Manual técnico.

11 / Recogida de datos adicional para el análisis sobre fiabilidad y validez.

¹³ Extraído de: Martínez Arias, Rosario.

¹⁴ Camilli y Shepard: "sesgo de un test es una fuente de invalidez o error sistemático que se refleja en cómo un test mide a los miembros de un grupo particular". El sesgo ha sido estudiado fundamentalmente en diferencias relacionadas con la etnia o el género. También, como la clase social, edad, religión, etc.

BIBLIOGRAFÍA

- Nunnally, Jum C. y Bernstein, Ira J. "Teoría Psicométrica". McGraw-Hill/Interamericana de México, 1995.
- Page, Mariano Á. "Elementos de Psicometría". Eudema, 1993.
- Esquivel, Juan M. "El diseño de las pruebas para medir logro académico: ¿Referencia a Normas o a Criterios? Costa Rica.
- Esquivel, Juan M. "Medición de la calidad de la enseñanza". Costa Rica, 1986.
- Martínez Arias, Rosario. "Psicometría: Teoría de los Test Psicológicos y Educativos". Editorial Síntesis, 1995.
- Meherens, W.A. "Medición basada en Normas y en Criterios". CECSA. México, 1982.
- UNAM. "La evaluación referida a un criterio: un modelo en el proceso de formación". Área de evaluación, 1976.
- IIFE – UNESCO. "Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa". Argentina, 2003.
- Sachs Adams, Georgia. "Medición y evaluación en educación, psicología y guidance". Barcelona, Herder, 1970.
- House, E. R. "Evaluación, ética y poder". Morata, 1994.
- Tenti Fanfani, Emilio (Comp.) "El rendimiento escolar en la Argentina". Losada, 2002.

Anexo 2

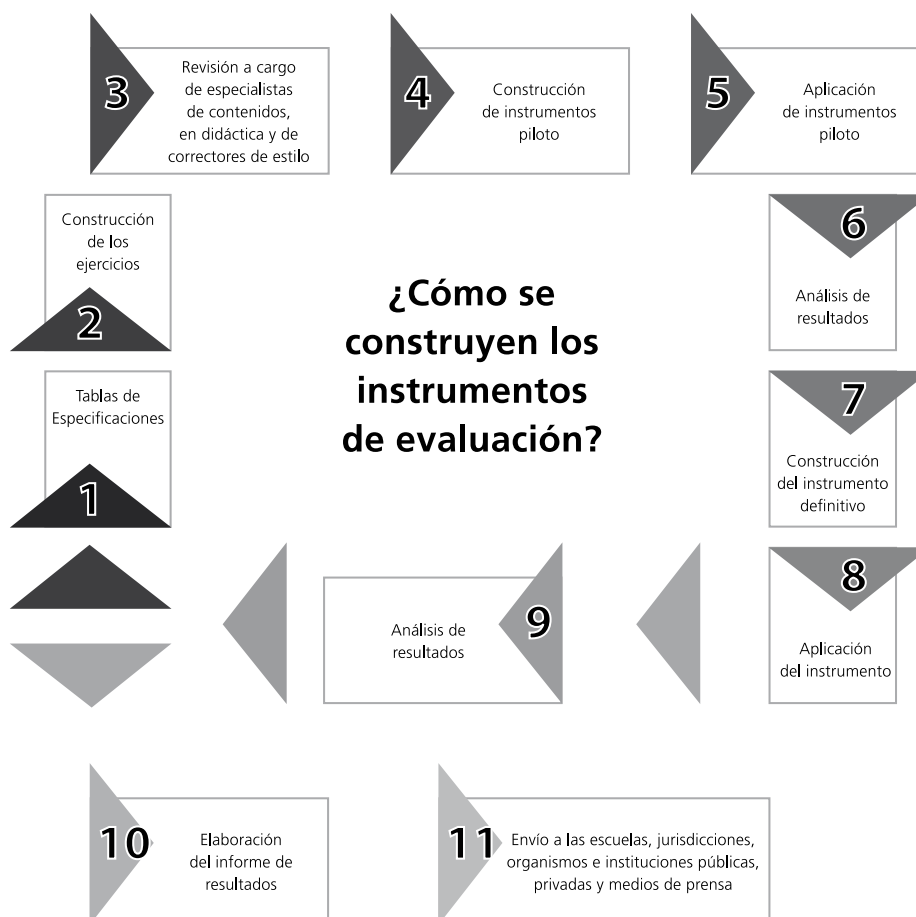


GRÁFICO 1 / Circuito metodológico y pedagógico de la elaboración y aplicación de los instrumentos de evaluación e interpretación y comunicación de los resultados.

Como parte de las acciones de evaluación, se elaboran pruebas que evalúan el dominio de contenidos y capacidades a fin de relevar información sobre el nivel de logros y dificultades de los alumnos de un determinado año en distintas áreas de conocimiento.

Las pruebas están conformadas por un conjunto de ejercicios cerrados, de respuesta fija de selección múltiple y por ejercicios de respuesta abierta, de desarrollo. Cada ejercicio de selección múltiple está constituido por un cuerpo y cuatro distractores. Sólo uno de ellos constituye la respuesta correcta.

Los ejercicios se diseñan, seleccionan y organizan de acuerdo con los contenidos y capacidades establecidos en las **tablas de especificaciones (PASO 1 DEL GRÁFICO 1)**.

Una de las formas usuales de las **tablas de especificaciones**

es la de un cuadro de doble entrada en el que se ingresan dos tipos de información: las capacidades y los contenidos a evaluar, que surgen de los **CBC** de la EGB y del Polimodal, de los diseños curriculares jurisdiccionales y de las opiniones de los expertos acerca de lo que se enseña en las escuelas.

Los contenidos y capacidades listados para cada área han surgido de acuerdos gestados entre las jurisdicciones a mediados de la década del '90. Los ejercicios tienen una distribución proporcional de valores al interior de cada tabla, justamente, por su peso en los **CBC** de la EGB y del Polimodal, en los diseños curriculares jurisdiccionales y según las opiniones de los expertos acerca de lo que se enseña en las escuelas. Cada cruce identifica una capacidad determinada en relación con un recorte del saber.

El conjunto final de contenidos y capacidades a ser evaluados cumple con las siguientes condiciones:

CONTENIDOS Y CAPACIDADES CONDICIONES

Son centrales desde el punto de vista de la disciplina.

Están presentes en los diseños curriculares de todas las jurisdicciones.

Tienen alta probabilidad de haber sido enseñados.

Tienen la posibilidad de ser evaluados en forma escrita.

Cuentan con la posibilidad de ser evaluados en forma masiva.

Por estas razones, las **tablas de especificaciones** son un instrumento indispensable para la construcción de las pruebas otorgándoles un alto nivel de validez y confiabilidad.

Al construir los ejercicios de las distintas áreas del saber (**PASO 2 DEL GRÁFICO 1**) se tienen en cuenta ciertos criterios:

CRITERIOS

La consigna de trabajo debe ser escrita en forma clara, breve y simple.

Deben tener vocabulario adecuado para los estudiantes.

Deben ser apropiados al nivel cognitivo de los estudiantes.

Deben tener un nivel apropiado de dificultad.

Deben tener una única respuesta correcta.

El cuerpo principal debe ser claro y poseer solo la información necesaria para la resolución del ejercicio.

La inclusión de gráficos y/o ilustraciones debe aportar información significativa que ayude a la comprensión del ejercicio y su resolución.

Cada uno de los distractores debe plantear una respuesta que, aunque errada, responda a "cierta lógica" del estudiante y/o presentar errores comunes o ideas falsas corrientes.

Los distractores deben ser similares en longitud, complejidad y estructura gramatical entre sí y con la respuesta correcta.

Una vez contruidos los ejercicios, son revisados por especialistas en contenidos que supervisan la corrección "científica" de los mismos; por pedagogos que analizan la pertinencia didáctica y por correctores de estilo que corrigen la sintaxis y la normativa (**PASO 3 DEL GRÁFICO I**).

Las pruebas se elaboran previendo, también, los diferentes niveles de dificultad de cada ejercicio. Estos niveles de dificultad, que en un principio responden a hipótesis de expertos, son probados con pequeños grupos de alumnos y en operativos piloto (**PASO 4 Y 5 DEL GRÁFICO I**).

A continuación, se presentan ejemplos de ejercicios con diferente nivel de dificultad:

Una vez probada en las muestras pilotos la calidad psicométrica de los ejercicios y sus niveles de dificultad, se integran a las pruebas en las proporciones establecidas en las tablas de especificaciones, tratando que los ejercicios de resolución fácil, mediana y difícil se encuentren distribuidos armónica y representativamente a lo largo de todo el instrumento para facilitar su resolución (**PASO 6 Y 7 DEL GRÁFICO I**). Los criterios de calidad estadística que se toman en cuenta para estas pruebas, a fin de seleccionar los ejercicios, son los siguientes:

A partir del análisis de los logros y de los errores cometidos con mayor frecuencia en los ejercicios, se plantean **recomendaciones metodológicas para la enseñanza** que pueden ayudar a superar estas dificultades (**PASO 9 Y 10 DEL GRÁFICO I**), que son enviadas a todas las escuelas, jurisdicciones y organismos e instituciones públicas y privadas, así como a la prensa para su difusión (**PASO 11 DEL GRÁFICO I**).

En suma, las pruebas de evaluación de la calidad se construyen para evaluar los niveles de rendimiento académico en distintos años del sistema educativo y en distintas áreas disciplinares; definiendo estos niveles en términos de los contenidos y de las capacidades que poseen los alumnos.

CRITERIOS DE CALIDAD PSICOMÉTRICA

Índice de discriminación: indica la capacidad del ejercicio para diferenciar entre los alumnos que obtienen alto rendimiento y los que obtienen bajo rendimiento.

Índice de dificultad: indica el porcentaje de alumnos que responde correctamente el ejercicio.

ADENDA

CAPACIDADES COGNITIVAS ONE 2009

MATEMÁTICA

En Matemática se evalúa una capacidad cognitiva general: la Resolución de Problemas.

A los efectos de la evaluación, se han considerado tres capacidades cognitivas específicas, incluidas en la resolución de problemas:

Capacidad Cognitiva general	Descripción
Resolución de problemas	Solución de situaciones nuevas para el alumno, en las que necesita usar los conocimientos matemáticos de que dispone. La resolución de problemas puede requerir de los estudiantes: reconocer, relacionar y utilizar información; determinar la pertinencia, suficiencia y consistencia de los datos; reconocer, utilizar y relacionar conceptos; utilizar, transferir, modificar y generar procedimientos; juzgar la razonabilidad y coherencia de las soluciones y justificar y argumentar sus acciones
Capacidades Cognitivas Específicas	Descripción
Reconocimiento de datos y conceptos	Capacidad cognitiva de identificar datos, hechos, conceptos, relaciones y propiedades matemáticas, expresados de manera directa y explícita en el enunciado.
Resolución de operaciones (mediante distintos procedimientos)	Reemplaza a la anterior Operar usando algoritmos. Consiste en la resolución de operaciones usando distintos procedimientos.
Resolución de situaciones en contextos intramatemáticos y/o de la vida cotidiana	Capacidad cognitiva de solucionar situaciones problemáticas contextualizadas, presentadas en contextos que van desde los intramatemáticos hasta los de la realidad cotidiana.
Comunicación en matemática	<p>Interpretar información:</p> <ul style="list-style-type: none"> ▪ Comprender enunciados, cuadros, gráficos ▪ Diferenciar datos de incógnitas ▪ Interpretar símbolos, consignas, informaciones ▪ Manejar el vocabulario de la matemática ▪ Traducir de una forma de representación a otra, de un tipo de lenguaje a otro <p>Expresión o emisión de procedimientos y resultados</p> <ul style="list-style-type: none"> ▪ Describir procedimientos de resolución utilizados ▪ Redactar correctamente la formulación de un resultado ▪ Describir las distintas etapas de una construcción geométrica ▪ Describir las distintas etapas de un cálculo ▪ Redactar una justificación, una argumentación ▪ Completar un cuadro, un gráfico ▪ Producir un texto, un cuadro, un gráfico, un dibujo ▪ Expresarse con un adecuado vocabulario matemático ▪ Formular un problema o situación problemática

LENGUA

CAPACIDADES COGNITIVAS: Operaciones mentales que el sujeto utiliza para establecer relaciones con y entre los objetos, situaciones y fenómenos. Se logran a través del proceso de enseñanza y del proceso de aprendizaje y cobran significado de acuerdo con la determinación de contenidos socialmente relevantes y altamente significativos, frente a los que se ponen en juego y a través de los cuales se desarrollan. En el área de Lengua, la capacidad general que evalúa nuestra prueba es la comprensión lectora. Dentro de ella se encuentran las siguientes capacidades:

EXTRAER: Localizar información en una o más partes de un texto.

Los lectores deben revisar, buscar, localizar y seleccionar la información. Deben cotejar la información proporcionada en la pregunta con información literal o similar en el texto y utilizarla para encontrar la nueva información solicitada.

INTERPRETAR: Reconstruir el significado global y local y hacer inferencias desde una o más partes de un texto.

Los lectores deben identificar, comparar, contrastar, integrar información con el propósito de construir el significado del texto.

EVALUAR Y REFLEXIONAR: Relacionar un texto con su propia experiencia, conocimientos e ideas.

Los lectores deben distanciarse del texto y considerarlo objetivamente. Deben utilizar conocimiento extra-textual (la propia experiencia, elementos proporcionados por la pregunta, conocimiento de mundo, conocimiento de la lengua, conocimiento de distintos géneros discursivos). Los lectores deben justificar su propio punto de vista.

Los contenidos específicos que se evalúan dentro de estas tres capacidades son:

EXTRAER	Información explícita. Secuencia de hechos o ideas. Resumen. Procedimientos textuales específicos de cada tipo textual (narración, explicación, argumentación).
INTERPRETAR	Tema, idea central, tesis. Relaciones textuales. Procedimientos de cohesión (referencia, paráfrasis, elipsis, conectores, etc.). Procedimientos textuales. Enunciación. Características de personajes. Vocabulario.
EVALUAR	Tipología textual. Géneros discursivos. Principios constructivos del texto literario. Tipos de narradores. Variedades y registros lingüísticos.

CIENCIAS SOCIALES

ANÁLISIS DE SITUACIONES: Capacidad cognitiva de reconocer relaciones y/o de seleccionar cursos de acción que requieren la aplicación de conceptos y/o principios y/o información previamente adquiridos.

COMUNICACIÓN: Capacidad cognitiva de interpretar la información y expresar un procedimiento o un resultado. Por razones metodológicas, los ítem que evalúan la comprensión o interpretación de la información se encuentran en la capacidad de Interpretación/

Exploración de la tabla de especificaciones.

Los ítems para evaluar la expresión de un procedimiento o de un resultado se incluyen, en la tabla, en la capacidad llamada Comunicación.

Las posibles alternativas para evaluar esta capacidad son las siguientes:

- Plantear una situación problemática.
- Describir la solución de una situación problemática.
- Redactar una fundamentación.
- Completar un cuadro, un mapa, etc.
- Producir un texto, un cuadro, un mapa, un gráfico, una tabla, etc.
- Expresarse con un adecuado vocabulario de la disciplina.
- Describir las distintas etapas de una construcción científica.

INTERPRETACIÓN/EXPLORACIÓN: Capacidad cognitiva de obtener y cruzar información proveniente de diferentes fuentes (textos, mapas, gráficos, etc.).

RECONOCIMIENTO DE CONCEPTOS: Capacidad cognitiva de identificar conceptos y principios por medio de ejemplos, casos, atributos o definiciones de los mismos o viceversa: identificar ejemplos, casos, atributos o definiciones de conceptos y principios dados.

RECONOCIMIENTO DE HECHOS: Capacidad cognitiva de identificar datos y/o hechos en un conjunto de información mediante la utilización de conocimientos que el alumno posee.

RECONOCIMIENTO DE VALORES: Capacidad cognitiva de elegir conductas de acuerdo con valores en situaciones vinculadas con la responsabilidad social e identificar valores implícitos o explícitos en discursos y prácticas.

CIENCIAS NATURALES

Capacidades Cognitivas	Definición
Reconocimiento de datos, hechos y conceptos	Incluye: la identificación de características y relaciones, la interpretación de datos y hechos: comparando, secuenciando, agrupando y clasificando según criterios explicitados. la aplicación de conceptos científicos.
Comunicación	Incluye: la interpretación, organización y traducción de información en distintos formatos (tablas, gráficos, diagramas, esquemas y símbolos), y la expresión de argumentos o conclusiones a partir de evidencias tales como datos experimentales.
Análisis de situación	Incluye: el análisis de los datos experimentales, su validez, la identificación de patrones y el reconocimiento de las variables involucradas, y la identificación, interpretación y relación de los distintos procesos de la investigación científica.

PROYECTO USO DE LA INFORMACIÓN 2009

Pautas para la presentación de Proyectos Jurisdiccionales 2009

A.- Antecedentes

En tres llamados distintos ('02-'03; '04 y '06-'07) se financiaron y asistieron técnicamente proyectos jurisdiccionales de uso de la información surgida de la evaluación de la calidad educativa.

En las revisiones que se han realizado de la propuesta, los responsables jurisdiccionales de evaluación destacaron tres puntos:

- Permitió que se implementaran propuestas que se venían formulando durante varios años pero que nunca habían contado con fondos para poder realizarse.
- Brindó la posibilidad de expandir el ciclo de evaluación de la calidad educativa a través de acciones de devolución y uso de la información obtenida.
- Respaldo acciones provinciales de evaluación para que puedan finalizarse y ser efectivas.

Como parte de la evaluación interna de la DiNIECE, se destacó:

- La importancia de la propuesta para impulsar el uso de información.
- La heterogeneidad de propuestas recibidas.
- La heterogeneidad en el nivel de implementación de los proyectos.

A lo largo de estos tres llamados se ha mejorado el seguimiento y la sistematización de la experiencia. Esto ha permitido contar con documentación sobre los proyectos de uso de información de los resultados de la evaluación de la calidad educativa en el país y la posibilidad de realizar un mejor seguimiento de los compromisos asumidos por las distintas partes.

B.- Pautas para la presentación de Proyectos Jurisdiccionales 2009

En convocatorias anteriores eran dos las condiciones que las Jurisdicciones debían respetar en sus proyectos: que se formularan de acuerdo al esquema de presentación propuesto por la DiNIECE y que fueran técnicamente viables.

Se agregan a los anteriores, los siguientes criterios:

- El Proyecto deberá estar orientado a mejorar el funcionamiento del sistema educativo, destinán-

dolo a funcionarios, supervisores, directivos y/o docentes.

- Los proyectos jurisdiccionales se centrarán en el uso de la información producida a partir del ONE '05.
- Las Jurisdicciones pueden utilizar, en forma complementaria de las bases del ONE '05, información surgida de estudios propios o evaluaciones provinciales.
- Estarán asesorados por la DiNIECE. Para el caso en que el proyecto contemple la participación de otra institución o profesionales externos, deberá contar previamente con el expreso acuerdo de la DiNIECE.
- La jurisdicción contará con un año (doce meses corridos) para ejecutar el proyecto a partir de recibir la transferencia de la DiNIECE. Vencido ese plazo, deberá nuevamente solicitar la aprobación del proyecto y la autorización para el uso de los fondos, explicando los motivos de la demora.

C.- Procedimiento administrativo

Las Jurisdicciones que lo deseen podrán presentar proyectos de uso de información, conforme a los criterios y esquemas que se postulan en el presente documento.

Al ser aprobado técnicamente por el Equipo Pedagógico de la DiNIECE, se remitirá, junto con un acta-acuerdo, a la Jurisdicción. Allí será firmado por la autoridad correspondiente y reenviado a la Dirección Nacional. Una vez que se dé entrada, y a la mayor brevedad posible, se realizará la transferencia de fondos.

La convocatoria está abierta hasta el 30/06/09. Es importante señalar que las transferencias pueden darse en distinto momentos del año. Sin embargo, la primera que realizará esta Dirección Nacional en el área de evaluación será en el mes de marzo de 2009. Para incluir al proyecto de uso de la información en esa partida éste deberá ser enviado antes de la segunda quincena de febrero.

D.- Monitoreo y seguimiento

Una vez aprobado el proyecto, la jurisdicción deberá presentar a la DiNIECE tres informes que documenten su ejecución:

- El primero será entre el primer y segundo trimestre de aprobado el proyecto. El mismo reflejará los procesos de inicio e implementación.
- El segundo describirá los procesos de ejecución y desarrollo.
- El tercero abarcará los procesos de finalización y evaluación del proyecto.

E.- Proyectos sugeridos

Respetando las condiciones y criterios fijados en puntos anteriores, cada jurisdicción podrá elaborar el proyecto que le parezca más adecuado a la realidad de su situación educativa y posibilidades. A modo de sugerencia, se señalan las siguientes propuestas:

- Jornadas de trabajo donde se presente y analice la información.
- Estas jornadas pueden estar dirigidas a funcionarios del sistema, a inspectores y/o directivos. Para ello, puede resultar útil la elaboración de material impreso donde se presenten los datos (Ejemplo: Proyecto presentado por la Provincia de La Rioja en la convocatoria 2006).
- Talleres de trabajo donde se presente la información, se la analice y se formulen recomendaciones o planes para mejorar la calidad educativa (Ejemplo: Proyecto presentado por la Provincia de Entre Ríos en la convocatoria 2006).
- Trabajo de devolución y formulación de planes de mejora en escuelas que tengan bajos resultados (Ejemplo: Proyecto presentado por la Provincia de Salta en la convocatoria 2006).

F.- Sobre el presupuesto y los fondos

Es importante que el presupuesto del proyecto esté bien definido.

Los rubros en que podrán utilizarse los fondos son:

- Asistencia técnica y profesional.
- Alquiler de Lugares y Equipos.
- Impresiones.
- Insumos de librería.
- Insumos informáticos (no equipos).
- Correo.

En cuanto al monto de los fondos que se soliciten, deben estar claramente detallados en el presupuesto, corresponderse con la propuesta que se realice y la realidad del sistema educativo jurisdiccional (tanto en dimensiones como en factibilidad de realización).

G.- Esquema para la presentación de proyecto

Los proyectos que se presenten utilizarán el esquema que se indica a continuación, similar al de otras convocatorias:

1. Marco General del Proyecto de Uso de la Información de Evaluación: refiere a una breve explicación de las razones que fundamentan la implementación del Proyecto, de acuerdo con las prioridades identificadas por la Jurisdicción, con respecto al uso de la información de evaluación. Se deberá explicitar por qué resulta necesaria su realización.

2. Grupo - meta: el Proyecto puede estar dirigido a diferentes actores del sistema educativo, de acuerdo con las necesidades que cada Jurisdicción identifique como prioritarias. La idea es caracterizar a los destinatarios del Proyecto, teniendo en cuenta:

- Cantidad aproximada de destinatarios.
- Nivel de decisión y acción de los mismos.
- Nivel/es educativo/s involucrado/s.

Por ejemplo: cantidad de supervisores de escuelas de EGB de la Provincia; cantidad de directivos y docentes de establecimientos de Nivel Polimodal de las regiones con rendimiento más bajo.

3. Objetivos: enunciar los logros que se pretenden alcanzar en relación con el uso de la información de evaluación con la implementación del Proyecto.

4. Estrategia general: supone explicitar el curso de acción a través del cual se prevé la consecución de los objetivos del Proyecto. Por ejemplo si se prevén encuentros de carácter presencial, una modalidad semipresencial o a distancia. Si los encuentros exigen el traslado de los destinatarios a algún punto de la Jurisdicción o el traslado del equipo responsable del Proyecto a distintos puntos de la misma. Si se elaborarán materiales de trabajo, etc.

5. Acciones: implica una mayor especificación de la estrategia general. Se deberán enunciar en forma secuencial las tareas y los responsables que las llevarán a cabo para la consecución de los objetivos.

6. Cronograma: especificar un listado de las acciones con las fechas tentativas de inicio y finalización.

7. Recursos:

- Materiales: folletos, copias, anillados, alguna bibliografía particular.
- Humanos: curriculum vitae de los responsables de la implementación del Proyecto.

8. Productos: refieren a los resultados a los que se llegará por haber realizado las acciones.

9. Presupuesto y Costos: implica explicitar el costo total y el de cada una de las acciones que se ejecutarán en el Proyecto. (VER PUNTO "F").

10. Seguimiento y Monitoreo:

- Informe inicial: Entre el 3er y el 6to mes de aprobado el proyecto. Reflejará los procesos de inicio e implementación.
- Informe de proceso: Entre el 6to y el 9no mes. Describirá los procesos de ejecución y desarrollo.
- Informe de cierre: Al finalizar la ejecución del proyecto. Abarcará los procesos de finalización y evaluación del proyecto.

11. Requerimiento a la DiNIECE:

- ASISTENCIA TÉCNICA: describir el tipo de asistencia que se requerirá y para qué momento. Por ejemplo: previa o durante la implementación del Proyecto.
- INFORMACIÓN DE EVALUACIÓN: bases de datos con los resultados de los ONE, diferentes tipos de procesamiento de la información, materiales de uso de la información de evaluación.
- OTRA (especificar).

H.- Apoyo desde la DiNIECE

Las jurisdicciones cuentan con el Equipo Nacional para asistirlos y ayudarlos tanto en el planteo de los proyectos como en su implementación. Por cualquier duda o consulta, pueden contactarse con Jorge Novello y/o Rafael del Campo al 011-4129-1486 int. 1486 o escribir a jnovello@me.gov.ar – rdelcampo@me.gov.ar.

BIBLIOGRAFÍA

- Criterios de evaluación ONE 2005. DINIECE, Ministerio de Educación, Ciencia y Tecnología de la Nación, 2005.
- Allen, David (compilador - 2000), La evaluación del aprendizaje de los estudiantes, Buenos Aires, Paidós.
- Barberá Gregori, Elena (1999), Evaluación de la enseñanza, evaluación del aprendizaje, Barcelona, edebé.
- Carrión Carranza, Carmen (2001), Valores y principios para evaluar la educación, México, Paidós.
- Cano García, E. (1999), Evaluación de la calidad educativa, Madrid, La Muralla.
- Landsheere, G. de (1999), El pilotaje de los sistemas educativos, Madrid, La Muralla – OEI.
- López, F. (1994) La gestión de la calidad en educación, Madrid, La Muralla.

SITIOGRAFÍA

- Elola, Nydia y Toranzos Lilia, (2000) Evaluación educativa. Una aproximación conceptual., Biblioteca Digital, OEI.

<http://www.campus-oei.org/calidad/luis2.pdf>

GLOSARIO

AGRUPAMIENTO DE ÍTEMS: Conjunto de ítems que evalúan una misma capacidad cognitiva o un mismo contenido. Este agrupamiento da la posibilidad de calcular un puntaje parcial de la prueba, de la cual forman parte.

ANÁLISIS DE SITUACIONES: Ver Capacidades Cognitivas.

ANCLAJES: Un conjunto de ítems comunes a pruebas diferentes que se aplican a grupos de alumnos también distintos y permiten la equiparación de los puntajes obtenidos.

APLICACIÓN CENSAL: Aplicación de una prueba de evaluación, a la totalidad de los alumnos de una cohorte de todo el país. Brinda información relevante sobre la población total.

APLICACIÓN MUESTRAL: Aplicación de una prueba de evaluación, a grupos de alumnos seleccionados de manera tal que sus características sean representativas de la población total. La selección de la muestra en nuestro caso, es aleatoria y estratificada (por jurisdicción, ámbito, régimen y modalidad).

AZAR: Ver Respuesta por Azar.

BANCO DE ÍTEMS: Un conjunto de reactivos (interrogantes, disparadores o ejercicios) organizados según el criterio lógico del área disciplinaria, curricular o por competencia, curso o edad de los alumnos, destinados a ser utilizados para construir futuros instrumentos de evaluación.

BLOQUE DE ÍTEMS: Conjunto de ítems de una prueba. Los bloques permiten una mayor cobertura de la diversidad de contenidos y capacidades. A partir de 2009 los ONE se estructuran en bloques.

CALIDAD (de la Educación): Una educación de calidad es aquella que logra que los alumnos realmente aprendan lo que se supone deben aprender, aquello que está establecido en los planes y programas curriculares al cabo de determinados ciclos del sistema educativo formal. Es trabajar atendiendo a los resultados e incrementar oportunidades de aprendizaje para todos con la permanente intención de ir disminuyendo las diferencias. Es aquella que brinda medios y orienta a los equipos de supervisión, docentes y alumnos para que se realice efectivamente el proceso de enseñanza y el proceso de aprendizaje, poniendo énfasis en las estrategias de gestión y didácticas.

CAPACIDADES COGNITIVAS: Operaciones mentales que el sujeto utiliza para establecer relaciones con y entre los objetos, situaciones y fenómenos. Se logran a través del proceso de enseñanza y del proceso de aprendizaje y cobran significado de acuerdo con la determinación de contenidos socialmente relevantes y altamente significativos, frente a los que se ponen en juego y a través de los cuales se desarrollan. En las tablas de especificaciones que enmarcan la aplicación de nuestros instrumentos de evaluación, quedaron establecidas hasta el momento capacidades cognitivas de acuerdo a la disciplina que se evalúe (Ver Adenda).

CBC: Contenidos Básicos Comunes. Conjunto de saberes relevantes que integran el proceso de enseñanza de todo el país. Son la matriz básica para un proyecto cultural nacional; matriz a partir de la cual, cada jurisdicción continua actualizando sus propios lineamientos o diseños curriculares y da paso, a

su vez, a diversos pero compatibles proyectos curriculares institucionales.

CIVED: Civic Education Study o Estudio Internacional de Educación Cívica. Prueba de evaluación internacional puesta en marcha por la IEA (International Association for the Evaluation of Educational Achievement) cuyo objetivo es evaluar los conocimientos y actitudes de los alumnos dentro de cuatro campos:

1. Democracia.
2. Identidad Nacional.
3. Cohesión Social y Diversidad.
4. Economía / Medios de Comunicación y Medio Ambiente.

En nuestro país esta prueba se aplica en 9° año de la EGB, o 2° año del nivel Secundario y en 3° año del nivel Polimodal o último año del nivel Secundario.

COMPRESIÓN LECTORA: Ver Capacidades Cognitivas.

CONFIABILIDAD: El grado de estabilidad de los puntajes obtenidos en mediciones sucesivas del mismo atributo o la consistencia de los puntajes arrojados por los distintos ítems que conforman la prueba.

CONTENIDOS: Conjunto de saberes o formas culturales cuya asimilación y apropiación por parte de los alumnos se considera esencial para su desarrollo personal y social. Son instrumentos para leer, analizar e interpretar la realidad. La explicitación de los contenidos, desde lo establecido por las diferentes disciplinas, permite determinar con mayor certeza de qué conocimiento debe apropiarse el alumno y cómo organizar las estrategias didácticas para implementar en los procesos de enseñanza y aprendizaje. Son seleccionados y organizados por el sistema educativo para ser enseñados y aprendidos en la escuela, sobre lo construido por la cultura vigente en un determinado momento histórico.

CUESTIONARIOS COMPLEMENTARIOS: También denominados como cuestionarios de contexto. Cuestionarios auto-administrados a los directivos, a los docentes y a los alumnos de las escuelas donde se llevan a cabo las pruebas de evaluación, a fin de recoger información relevante acerca de las condiciones materiales, institucionales y sociales en las cuales, la comunidad educativa desarrolla sus respectivas tareas.

DIFICULTAD DE UN ÍTEM: Medida que indica cuán difícil resultó un ítem para los estudiantes que lo resolvieron. Sus valores oscilan entre menos infinito y más infinito en la escala logit usada en el modelo Rasch. En términos más prácticos, los ítems asumen valores entre -3 y +3. Valores positivos y altos indican alta dificultad y valores negativos indican baja dificultad.

DISTRACTORES: Cualquiera de las opciones o respuestas erróneas posibles, que conforman los ejercicios de selección múltiple o de respuesta fija.

EQUIDAD: Calidad para todos. Es reconocer puntos de partida distintos y puntos de llegada similares, a partir de una visión crítica y objetiva de datos confiables. Una escuela es equitativa cuando posee la capacidad de compensar mejor las desigualdades sociales, debilitando el condicionamiento del origen social de los alumnos con respecto a sus logros de aprendizaje.

EQUIPARACIÓN: Proceso estadístico que se utiliza para ajustar puntuaciones de dos o más pruebas de modo que sean intercambiables y comparables, es decir que se pueda establecer una correspondencia entre sus puntuaciones, de tal modo que las puntuaciones de cualquiera de ellas puedan expresarse en términos de las de las otras.

ESCALAMIENTO: El desarrollo de reglas sistemáticas y de unidades significativas de medida para cuantificar las observaciones empíricas. Una escala de medida se establece cuando se define el conjunto de valores posibles que pueden asignarse y se establece la regla de asignación que establece la correspondencia entre el sistema empírico y el sistema numérico. Proceso de transformación de las respuestas en puntuaciones.

EVALUACIÓN: Proceso sistemático de relevamiento de datos, incorporado al sistema de acciones educativas, que permite obtener información válida y confiable para formular juicios de valor acerca de una realidad, en este caso nos referimos al campo educativo. Estos juicios pueden ser utilizados en la toma de decisiones con el objeto de mejorar las acciones educativas valoradas.

FACTORES ASOCIADOS: Variables sociales, culturales y económicas que inciden en el rendimiento académico de los alumnos.

INDICADOR: Parámetro que permite cuantificar e inferir el valor y la existencia o inexistencia de una variable. Da cuenta de cómo vamos a traducir en el ámbito empírico nuestra definición conceptual de la variable. Sirve de puente entre las conceptualizaciones y la realidad empírica.

INTERPRETACIÓN Y EXPLORACIÓN:
Ver Capacidades Cognitivas.

INSTRUMENTO PILOTO: Prueba de evaluación denominada "piloto" que consiste en una instancia previa a la construcción del instrumento definitivo en la que se administra un conjunto de ítem a una muestra de sujetos, a fin de obtener información sobre la calidad y pertinencia de dichos ítem para evaluar los contenidos y las capacidades previamente establecidos en las tablas de especificaciones.

INFORME DE RESULTADOS: Publicación destinada a la difusión de la información relevada sobre los aprendizajes de los alumnos y de los datos vinculados con su historia académica, las expectativas de los docentes respecto de éstos, las formas de organización del trabajo docente y de la escuela en su conjunto. Información que se recoge mediante la aplicación de pruebas estandarizadas y de cuestionarios complementarios, respectivamente.

INTERPRETACIÓN Y EXPLORACIÓN:
Ver Capacidades Cognitivas.

ÍTEM: Mínima parte de un test o prueba que recibe puntaje. Los formatos de ítem más utilizados para evaluar rendimiento académico son los de elección múltiple o de respuesta fija, en los cuales el alumno debe elegir la respuesta correcta de entre un conjunto limitado de respuestas posibles y los de construcción o ítem abierto, en los cuales el alumno debe construir la respuesta.

NAP: Núcleos de Aprendizaje Prioritarios. Constituyen un conjunto de saberes comunes que deben estar al alcance de todos los niños y las niñas del país, de manera tal que nos permita compartir el mundo a todos los argentinos, y reafirmar, desde el Estado, el derecho y la oportunidad de todos a acceder a nuestra cultura. (El acuerdo alcanzado en el Consejo Federal de Cultura y Educación, entre el Ministerio nacional, las provincias y la Ciudad de Buenos Aires, permitió establecer los Núcleos de Aprendizajes Prioritarios, conformándose una base común para

la enseñanza en todo el país).

NIVELES DE DESEMPEÑO: Categorías de tareas que permiten identificar grupos de estudiantes con niveles similares de rendimiento frente a la prueba. Se establecen fundamentalmente con el propósito de facilitar la comunicación de lo que pueden hacer los estudiantes. Se determinan combinando los hallazgos de los ONE y los criterios conceptuales curriculares y pedagógicos.

NIVELES DE LOGRO: Ver Niveles de Desempeño.

OMISIÓN: Ausencia total de respuesta en un ejercicio. El índice de omisión es el porcentaje de ejercicios no respondidos sobre el total de la población de alumnos evaluados.

ONE (Operativo Nacional de Evaluación): Acciones de evaluación instrumentadas a partir del nuevo marco legal otorgado por la sanción de la Ley Federal de Educación, con el objetivo de suministrar información válida y confiable sobre qué y cuánto aprenden los alumnos durante su permanencia en el sistema educativo formal y acerca de cuáles son los factores asociados a ese aprendizaje.

OPERACIÓN USANDO ALGORITMOS:
Ver Capacidades Cognitivas.

PIRLS: Progress in International Reading Literacy Study o Estudio internacional sobre el progreso de la alfabetización lectora. Prueba de evaluación internacional puesta en marcha por la IEA (International Association for the Evaluation of Educational Achievement), cuyo objetivo es evaluar tanto los procesos que se ponen en juego para la comprensión como los propósitos de lectura. En nuestro país se aplica en 4º año de la E.G.B. (Ver documento " Pruebas Nacionales e Internacionales para 2001").

PISA: Programme for International Student Assessment o Programa para la evaluación internacional de estudiantes. Prueba de evaluación internacional o programa creado por la OECD (organización para la Cooperación y el Desarrollo Económico) que evalúa tres áreas:

1. Prueba de Alfabetización Lectora estructurada sobre la base de tres dimensiones: Procesos o tareas de lectura - Contenidos o tipos de textos - Contextos o propósitos del texto.
2. Prueba de Alfabetización Científica estructurada sobre la base de tres dimensiones: Procesos Científicos - Conceptos Científicos - Situaciones y Contextos.
3. Prueba de Alfabetización Matemática estructurada sobre la base de tres dimensiones: Las Grandes Ideas (contenidos) - Competencias matemáticas o procesos de matematización - Situaciones y contextos.

El objetivo del estudio PISA 2000 es elaborar indicadores del alcance de la preparación de los alumnos de 15 años de los sistemas educativos de los países participantes, ya que ésta es la edad en la que la mayoría de los alumnos de los países miembros de OECD terminan su escolaridad obligatoria. La evaluación no se limita a lo que los alumnos han aprendido sino que pone especial interés en el uso que hacen de estos aprendizajes para jugar un papel comprometido como ciudadanos.

PRODUCCIÓN ESCRITA: Ver Capacidades Cognitivas.

PRUEBA PILOTO: Ver Instrumento Piloto.

PRUEBA REFERIDA A CRITERIOS: Los instrumentos referidos a criterios están enfocados a determinar si un alumno ha logrado un nivel predeterminado o criterio prefijado, en una disciplina dada. Proveen información precisa sobre logros respecto de ese criterio (estándares o metas educativas).

PRUEBA REFERIDA A NORMAS: Los instrumentos referidos a normas comparan al estudiante con la norma de su grupo.

Las mediciones normativas reportan cuál es la posición relativa de un sujeto con respecto de la norma de su grupo.

RECOMENDACIONES METODOLÓGICAS: Publicaciones y videos destinados a los docentes, cuyo contenido surge del análisis y de la interpretación de los resultados alcanzados por los alumnos en cada uno de los Operativos Nacionales de Evaluación. En este material se plantean propuestas, tanto teóricas como metodológicas para la enseñanza de aquellos contenidos y capacidades, que en forma recurrente presentan dificultad en los alumnos.

RECONOCIMIENTO DE CONCEPTOS Y PRINCIPIOS: Ver Capacidades cognitivas.

RECONOCIMIENTO DE HECHOS: Ver Capacidades cognitivas.

RECONOCIMIENTO DE VALORES: Ver Capacidades cognitivas.

REFLEXIÓN SOBRE LOS HECHOS DEL LENGUAJE: Ver Capacidades cognitivas.

RESOLUCIÓN DE PROBLEMAS: Ver Capacidades cognitivas.

RESPUESTA POR AZAR: Respuesta que los alumnos dan fortuitamente, sin dominar el contenido y capacidad específicos evaluados en el ítem.

SERCE: Segundo Estudio Regional Comparativo y Explorativo (SERCE 2006). El SERCE es un Estudio de Evaluación, producto de una construcción colectiva y participativa de los países latinoamericanos que conforman la red del Laboratorio Latinoamericano de Evaluación de Calidad Educativa (LLECE).

SISTEMA NACIONAL DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN: dispositivo administrativo para la conducción del sistema educativo. Su función es ofrecer información sobre los resultados de la educación, el funcionamiento de distintos niveles del sistema, las condiciones que afectan los procesos en las distintas instancias y el aporte de distintos actores. Esto con el propósito de definir las políticas educativas, orientar el desarrollo de planes de mejoramiento y rendir cuentas sobre la respuesta del sistema a las demandas de la sociedad.

TABLA DE ESPECIFICACIONES: En el caso de nuestras pruebas de evaluación es un cuadro de doble entrada en el cual se ingresan dos tipos de información: las capacidades y los contenidos a evaluar. Cada cruce identifica un desempeño, es decir, una capacidad determinada en relación con contenidos de cada área de conocimiento.

TEORÍA CLÁSICA DE LAS PRUEBAS (TCP): ver Teoría Clásica de los Test (TCT).

TEORÍA CLÁSICA DE LOS TEST (TCT): Teoría basada en el Modelo Lineal de la Regresión con dos variables: teoría psicométrica que considera como hipótesis fundamental, que el puntaje de un sujeto en un test (observable) es función de dos componentes: el puntaje verdadero (inobservable) y el puntaje de error.

TEORÍA DE LA RESPUESTA AL ÍTEM (TRI): Teoría basada en la modelización de las probabilidades de respuestas correctas a un ítem según las funciones de distribución normal y logística. Intenta dar una fundamentación probabilística al problema de la medición de constructos inobservables. Considera al ítem como unidad básica del test. Sus modelos son funciones matemáticas que relacionan la probabilidad de una respuesta particular a un ítem, con la aptitud general del sujeto.

TEORÍA DEL RASGO LATENTE: ver Teoría de la Respuesta al Ítem (TRI).

TIMSS: Third International Mathematics and Science Study o Tercer Estudio Internacional de Matemática y Ciencias. Prueba de evaluación internacional, puesta en marcha por la IEA (International Association for the Evaluation of Educational Achievement). Su objetivo es evaluar el nivel de rendimiento de los alumnos, de manera conjunta en Matemática y Ciencias. El núcleo central del estudio es la población de alumnos de 13 años. (Ver documento "Pruebas Nacionales e Internacionales para 2001").

VALIDEZ: El grado en que un test o prueba mide lo que dice medir o evalúa lo que pretende evaluar. Brinda información válida con respecto al atributo que mide o evalúa el test.

BIBLIOGRAFÍA

Coll, César. Aprendizaje escolar y construcción del conocimiento, Barcelona, 1990.

Coll, César; Pozo, J.I.; Sarabia, B; Valls, E. Los contenidos en la Reforma. Enseñanza y aprendizaje de conceptos, procedimientos y actitudes, Madrid, 1992.

Cortada de Kohan, Nuria. Teorías Psicométricas y Construcción de Tests, Buenos Aires, 1999.

Instituto para el Desarrollo de la Calidad Educativa. Documento Pruebas Nacionales e Internacionales para 2001,

Hambleton, R.K; Swaminathan, H; Jane Rogers, H. Fundamentals of Item Response Theory, United States of América, 1991.

Jacob, 1994; tomado de Cea D'Ancona, Ma. Ángeles. Metodología Cuantitativa. Estrategias y Técnicas de Investigación Social, Madrid, 1996.

Martinez Arias, Rosario. Psicometría: Teoría de los Tests Psicológicos y Educativos, Madrid, 1995.

Ministerio de Cultura y Educación de la Nación. Manual de Estrategias para el uso e incorporación de la información de la Evaluación, Buenos Aires, 1997.

Ministerio de Cultura y Educación de la Nación. Recomendaciones Metodológicas para la Enseñanza - 3er Operativo Nacional de Evaluación (1995), Buenos Aires, 1997.

Ministerio de Cultura y Educación de la Nación. Contenidos Básicos Comunes para la Educación General Básica, Buenos Aires, 1995.

Ministerio de Cultura y Educación de la Nación. Núcleos de Aprendizaje Prioritarios para la Educación General Básica, Buenos Aires, 2004 - 2006.

Ministerio de Cultura y Educación de la Nación. Propuesta de Tablas de Especificaciones, Buenos Aires, 1999.

Vianna, Heraldo Marelim. Termos Técnicos em medidas educacionais, São Paulo, 1981.

