

LAS PRUEBAS MASIVAS

Análisis de sus diferencias técnicas

ÁNGEL DÍAZ BARRIGA

Resumen:

En los últimos diez años se ha dado un significativo impulso en México a la aplicación de pruebas masivas, técnicamente llamadas a gran escala. Si bien desde mediados de los setenta han operado estos instrumentos, hasta ahora los resultados se han difundido entre la sociedad. Sin embargo, a la fecha no hay un análisis técnico del significado de los elementos que subyacen en su construcción. En este estudio se analizan, desde tres perspectivas, dos pruebas: el EXANI I, que se aplica para ingresar al bachillerato, y la de PISA, para estudiantes de 15/16 años. El autor examina las teorías: del test, que subyace en cada instrumento; de aprendizaje, con la que se construyen los reactivos; y del contenido, que se encuentra detrás de los reactivos de matemáticas y lenguaje. Concluye que las pruebas mexicanas responden conceptualmente a una generación anterior a las que está empleando la OCDE.

Abstract:

During the past ten years, significant impetus has been given in Mexico to the use of tests for the masses, known technically as large-scale. Although such testing instruments have been in operation since the mid 1970s, only recently have the results been made known to the general population. To date, no technical analysis has been made of the meaning of the underlying elements in their construction. This study analyzes, from three perspectives, two tests: EXANI-I, the high school admissions test; and the PISA test for students aged fifteen and sixteen. The author examines the theories: of the test, which underlies each instrument; of learning, used to construct responses; and of the content, found behind the responses of mathematics and language. The conclusion is that Mexican tests are conceptually one generation behind the tests used by OECD.

Palabras clave: pruebas de admisión, pruebas de aprovechamiento, análisis comparativo, tests y teorías del aprendizaje y del contenido, México.

Key words: admissions tests, achievement tests, comparative analysis, tests and theories of learning and content, Mexico.

Ángel Díaz Barriga es investigador del Centro de Estudios sobre la Universidad de la UNAM. Edificio de la Unidad Bibliográfica, lado norte del Centro Cultural Universitario, Ciudad Universitaria, Coyoacán, CP 04510, México, DF, CE: adbc@servidor.unam.mx

La evaluación educativa se ha desarrollado recientemente como una disciplina compleja. Lejos estamos de la perspectiva que la circunscribía al aprendizaje¹ y, aun en este ámbito, va conformando nuevas técnicas y diferentes procesos de trabajo.

Mundial y nacionalmente los responsables de la gestión de los sistemas educativos se encuentran a la búsqueda de estrategias que permitan mostrar que hay mejores indicadores del logro educativo. En este contexto, las pruebas masivas permiten realizar comparaciones de mejoras de aprendizaje que se manifiestan en una expresión numérica, con la finalidad de realizar comparaciones de resultados al interior de un país o entre varios de ellos. En este caso la técnica –los instrumentos técnicos de evaluación– por una parte posibilitan la realización de estas comparaciones y, simultáneamente, ocultan² el problema de fondo: el desarrollo de procesos de aprendizaje en los estudiantes. La dinámica a la que responden este tipo de instrumentos guarda relación con varios elementos:

- 1) En primer término, el crecimiento de la matrícula de los sistemas educativos: en la educación formal participan actualmente en la primaria 360 millones de estudiantes y en la secundaria 247 millones de alumnos a nivel mundial³ (ver cuadro 1).
- 2) En segundo lugar, el interés permanente de la escuela del Estado nacional para “comparar” los sistemas educativos, no podemos perder de vista que la educación comparada, como subdisciplina del campo, se empezó a conformar a finales del siglo XIX –a partir de las legislaciones que establecían la obligatoriedad del nivel básico de diversos países de Europa–, pues permitía analizar los contenidos que se estudiaban en cada país;⁴ actualmente las pruebas nacionales e internacionales son los instrumentos que permiten realizar esa comparación, ya no de los contenidos que forman el plan de estudios, sino de los resultados del esfuerzo educativo.
- 3) Finalmente, las pruebas masivas tienen un papel en las recientes discusiones acerca de la calidad de la educación que reciben los estudiantes; por medio de ellas se busca tanto reconocer los niveles de aprendizaje que están obteniendo, como comparar la calidad de los sistemas educativos o de las instituciones escolares a través de los resultados de los estudiantes en dichos exámenes.

CUADRO 1

Total de estudiantes en educación primaria y secundaria 2001-2002

Nivel educativo	Mundial*	América Latina*	México**
Primaria	360 995 799	47 047 094	14 843 381
Secundaria	247 315 090	44 157 731	8 424 294

* Datos tomados de *Compendio mundial de la Educación 2004. Comparación de estadísticas de educación en el mundo* (http://www.uis.unesco.org/TEMPLATE/pdf/ged/2004/GED2004_SP.pdf). La UNESCO clasifica la educación media (secundaria y bachillerato para México) como enseñanza secundaria.

** Datos tomados de *Estadísticas básicas del sistema educativo nacional* (<http://www.sep.gob.mx/work/appsite/princip2002/publicacion2003.pdf>).

La aplicación de estas pruebas masivas es el resultado tanto del desarrollo de la teoría de los test, como de la aplicación de las tecnologías de la información para el procesamiento de los datos. Lo que ha permitido realizar un conjunto de tareas técnicas referidas no sólo a calificar y obtener medias de cada prueba, sino una serie de procesos estadísticos para determinar su confiabilidad, su índice de dificultad, su poder de discriminación. Esto es, para validar la medición del instrumento que será objeto de aplicación.

Asimismo, conviene hacer una puntualización, ya que el manejo técnico que actualmente se realiza de los test se encuentra lejano de aquél con el que se formaba a los profesionales de la psicología y la educación hace apenas unos veinte años, cuando sólo se trabajaba con su teoría clásica, cuyo manejo se realizaba apoyado en elementos de la estadística descriptiva. En la actualidad, el empleo de la llamada teoría de la respuesta al ítem —con el apoyo en modelos matemáticos complejos basados en la determinación del coeficiente de regresión— constituyen un elemento fundamental que, junto con la informática, han modificado el manejo de los instrumentos de medición del aprendizaje, potencializando su uso.

En cierto sentido podemos afirmar que, desde el punto de vista técnico, nos encontramos en una nueva fase de la teoría del test. En este punto es importante plantear que identificamos un desfase en la mayoría de los planes de estudio sobre la incorporación de estas nuevas perspectivas en la formación de las futuras generaciones en el campo de la evaluación.

En segundo término es importante reconocer que la aplicación de pruebas con la finalidad de comparar países tiene cerca de cincuenta años en proceso de construcción. Así, las primeras experiencias de aplicación de pruebas masivas las encontramos en los años cincuenta del siglo pasado. En un proceso en que se fueron conformando diversas agencias internacionales ocupadas de esta tarea. No necesariamente éstas forman parte de los organismos internacionales, aunque recientemente algunos de ellos –la UNESCO y la OCDE– se han dedicado a esta labor. La más antigua es la Asociación Internacional de Evaluación (IEA) que, desde los años cincuenta, inició sus actividades en este sentido, al aplicar exámenes principalmente en diversos países europeos. Espacio donde se conformó lo que actualmente se conoce como prueba TIMSS (*Third International Mathematic and Science Study*), que se utiliza en México desde principios de la década de los años noventa, pero cuyos antecedentes en Europa se encuentran en el *First International*, desarrollado a mediados de los sesenta (cuadro 2).

A partir de los últimos cincuenta años, se ha desarrollado una serie de pruebas basadas en la teoría del test para ser aplicadas a un número grande de sujetos. Técnicamente se les denomina a gran escala, sin embargo, por la evolución y cantidad de sujetos que son objeto de su aplicación también pueden denominarse masivas. Este tipo de exámenes se puede considerar un fenómeno relativamente reciente en el campo evaluativo. Su evolución ha dado paso al establecimiento de diversas agencias nacionales e internacionales para establecer, por lo menos, dos generaciones de este tipo de exámenes. Lejos de lo que suele pensar la mayoría de las personas, cada prueba procede de una serie de presupuestos conceptuales distintos, dicho de otra manera, cada prueba mide un atributo diferente.

CUADRO 2

Principales agencias internacionales que han desarrollado pruebas masivas

-
- International Association for the Evaluation of Educational Achievement (IEA), inicia actividades en la década de los años cincuenta.
 - TIMSS (Third International Mathematic and Science Study). El First Study fue aplicado en 1964.
 - Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE-OREALC-UNESCO), inicia actividades en 1994.
 - PISA (Programme for International Student Assessment), OCDE, inicia sus actividades en 2000.
-

Es como si un automóvil fuera al mismo tiempo evaluado por un sector que valora su funcionalidad como coche familiar, por otro que mide su posibilidad para emplearlo como de carreras y por un tercer grupo que lo evalúa como auto de carga. Los resultados de la evaluación, aun cuando se expresen en puntajes, no se refieren al mismo atributo. Igual sucede con los distintos exámenes masivos que se han elaborado en el ámbito educativo, lo que nos obliga a no generalizar sus resultados.

Los objetivos de este trabajo son mostrar la existencia de diferencias significativas en los presupuestos conceptuales y técnicos, en la perspectiva del aprendizaje y en la visión del contenido entre dos pruebas: una nacional, la que se aplica como examen de ingreso al bachillerato (EXANI-I) a estudiantes que egresan de la secundaria, y una internacional, la que recientemente aplicó la OCDE en el marco de su *Programme for International Student Assessment*, mejor conocido como prueba PISA. La finalidad de nuestro estudio es doble, por una parte mostrar la existencia de una diferencia sustantiva entre ambos instrumentos de evaluación, por la otra contribuir al campo de la evaluación educativa mostrando nuevos desarrollos conceptuales que permitan realizar una serie de precisiones sobre tales instrumentos, así como matizar el manejo de los resultados que se obtienen en los mismos.

El trabajo de comparación demanda realizar algunos acotamientos. Aunque el objetivo de ambas pruebas es diferente, puesto que la de PISA tiene como meta comparar hasta dónde el aprendizaje que han tenido jóvenes de 15 años en el sistema educativo les ha permitido desarrollar estrategias y habilidades para la vida. Mientras que el EXANI-I es una prueba para seleccionar a los estudiantes para su ingreso al bachillerato de acuerdo, no con el nivel académico ideal que deben mostrar para lograr su ingreso sino, fundamentalmente, en función del cupo que tienen las instituciones de enseñanza media superior.⁵

Ambas pruebas se aplican a jóvenes de 15 años, elemento que nos induce a considerar que, aun reconociendo la diferencia de su finalidad, es posible realizar una comparación entre ambos instrumentos. La de PISA realizó un desarrollo conceptual tanto sobre su perspectiva del papel del aprendizaje en la vida de los individuos, como sobre el mecanismo que puede ser empleado para delimitar los contenidos objeto de indagación. Por su parte, el EXANI-I toma como referencia el plan de estudios de la primaria y la secundaria, para destacar aquellos contenidos que permitirán un mejor desempeño en la educación media superior.

Por esta razón optamos por establecer tres ejes de análisis que permitieran valorar el papel educativo que subyace en ambas pruebas. Tales ejes son: *a*) la noción de contenido en la que se fundamenta la construcción de los reactivos; *b*) las concepciones de aprendizaje que subyacen en cada instrumento; y *c*) su concepción de evaluación, en la que incluimos tanto los aspectos técnicos para la construcción de cada reactivo como el manejo que realizan de la información que se obtiene de la misma.

La conformación y evolución del sistema de exámenes masivos en México

En el caso mexicano estas pruebas masivas tienen una larga experiencia. La teoría del test, aplicada a la medición del rendimiento escolar, es una práctica que se fue conformando desde los años cincuenta; a principios de la década de los sesenta, la Universidad Nacional Autónoma de México la empleó como instrumento de selección de estudiantes que aspiraban ingresar a la licenciatura en 1962 y al bachillerato desde 1964.⁶ A principios de los años setenta el gobierno federal establece la Dirección General de Evaluación (SEP, 2002) aplicando inicialmente, desde 1974, diversos exámenes a estudiantes de secundaria y, desde 1984, pruebas generales de educación básica y normal.

Múltiples exámenes nacionales se aplicaron desde esa fecha a los estudiantes de las escuelas mexicanas, quizá el más conocido sea el de sexto grado denominado Instrumento de diagnóstico de alumnos de nuevo ingreso a la secundaria, conocido como IDANIS⁷ para determinar los conocimientos y habilidades que han desarrollado en el campo de las matemáticas y el lenguaje.

En una segunda etapa, en el marco del conjunto de reformas educativas mundiales, en los años noventa –la década que podríamos denominar de las reformas y de la evaluación de la educación– se crearon otras instituciones para la aplicación de pruebas masivas, tal es el caso del Centro Nacional de Evaluación (CENEVAL), establecido en 1994 con la responsabilidad de impulsar el establecimiento de un sistema de exámenes de ingreso y egreso de la educación media superior y superior⁸ y, recientemente en 2002, el Instituto Nacional de Evaluación Educativa (INEE), este último con la finalidad de desarrollar un sistema de pruebas que permitan comparar los resultados de los estudiantes de los diversos niveles del sistema educativo.

A diferencia de la etapa anterior en la que la autoridad educativa no autorizaba publicar los resultados que obtenían los estudiantes en los exámenes,⁹ por lo que la Dirección General de Evaluación se limitaba sólo a elaborar informes con carácter restringido para las autoridades educativas.¹⁰ La fase actual de evaluación es más abierta, lo que no necesariamente significa que se cumpla con la función pedagógica al promover una tarea formativa de la evaluación, ni con la función de retroalimentación inherente a la acción de evaluar.

Recientemente los aspirantes a la educación media superior y superior reciben no sólo un puntaje de los resultados, sino las áreas de conocimiento en la que obtuvieron los mejores y los más deficientes. De igual forma, instituciones como el INEE y el CENEVAL empiezan a publicar un conjunto de datos que van un poco más allá de los puntajes obtenidos.

Es importante enfatizar que mientras inicialmente esta información no sólo era escasa y en general periodística, en los últimos años la nueva gestión del CENEVAL está impulsando una serie de publicaciones propias, así como reportes periódicos que reflejan otra visión de la evaluación fundamentalmente del manejo de sus resultados.¹¹ Sin embargo, las situaciones contractuales que subyacen en varias de las acciones de evaluación del Centro obligan a un manejo circunscrito de tales informes, cumpliendo con la caracterización que Stufflebeam y Shinkfield (1987) hacen de las condiciones de la evaluación determinada por contratos. La tensión entre confidencialidad y manejo de la información aparece con claridad en aras de lograr la autosuficiencia económica del Centro.

A partir de la creación del INEE, en 2002, se está avanzando en el establecimiento de un sistema de exámenes nacionales, en el análisis de los que aplicaba la Dirección General de Evaluación,¹² así como de la información que de ellos se derivaba, al tiempo que se apoyan las pruebas internacionales y se ofrece al público una serie de documentos especializados en evaluación.¹³ Las nuevas tecnologías de la información y, en particular internet, también permiten en este momento tener acceso a una serie de documentos elaborados por el Instituto.

De igual forma debemos reconocer que esta información no permite realizar cabalmente la tarea formativa y de retroalimentación. El docente es visualizado por ambas instancias como un consumidor tangencial de esta información y no como un importante destinatario. Más aún en una clara distorsión del sentido de la evaluación –ante la sorpresa del mundo

académico—, el INEE ha establecido diversos talleres de evaluación para periodistas. Así lo consigna en su informe de actividades bajo los siguientes argumentos:

[...] para fomentar el desarrollo de una cultura de la evaluación en la sociedad mexicana, y además de la atención cotidiana a los medios, desde febrero de 2003, el área inició un curso de actualización para periodistas, con sesiones mensuales, en colaboración con la Organización de Estados Iberoamericanos y Observatorio Ciudadano de la Educación. Se han realizado seis sesiones, con asistencia promedio de 50 personas. Se prepara un prontuario de información para el periodista educativo, folletos de difusión dirigidos a padres de familia y maestros, así como el programa de comunicación de mediano y largo plazos del INEE.¹⁴

Todo ello sólo contribuye a incrementar el carácter judicativo, esto es, la emisión de un juicio, de una calificación, en detrimento de la retroalimentación. No basta, como se puede comprobar ampliamente con lo que ha acontecido en los últimos años, con promulgar información procedente de la evaluación. Ello, incluso, puede incrementar una sospecha de que ésta se realiza contra el docente, para someter su actividad a un escarnio público, sin analizar las múltiples condiciones y diversos factores que afectan el proceso de aprendizaje escolar. Esta posición se apoya en una perspectiva de los resultados de la investigación educativa que consideran que el docente es el principal obstáculo para el cambio en la educación (Anagnostopoulos, 2005) y, al mismo tiempo, omiten evidencia significativa, también resultado de la investigación en educación, donde se muestra que el cambio de la educación se realiza a partir de los docentes y que estos profesionales se pueden comprometer ampliamente con ello (Teobaldo, Bertoni y Poggi, 1995).

Así, por ejemplo, un profesor de matemáticas de secundaria nunca sabe cuáles son los temas del plan de estudios donde los alumnos tienen mayores deficiencias. Mientras esto no se pueda informar, los profesores de grupo —pues en el aula es donde realmente se puede mejorar la educación— no tendrán condiciones para realizar el cambio que se espera en pro de la calidad. Es como si un mecánico le dijera al chofer del automóvil “su carro tiene problemas por eso se descompone”, un electricista al dueño de una casa “el sistema eléctrico está mal”. Mientras no le indique el sentido de la

falla no se puede hacer nada. De nada sirve, en un sentido pedagógico, que se informe a la sociedad que los alumnos mexicanos ocupan el lugar 32 en la prueba PISA o que tienen un promedio de conocimientos de 4.5 en matemáticas de secundaria. Por eso, en otras ocasiones, hemos afirmado que el sistema de evaluación mexicano tiene más un carácter judicativo que pedagógico.

La aplicación de pruebas internacionales en el escenario nacional

Los exámenes internacionales para estudiantes mexicanos inician a mediados de la década de los noventa participando en los elaborados por el TIMSS y del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa. Así, en 1998, el Laboratorio de la Calidad de la Educación para América Latina y el Caribe aplicó baterías de exámenes de español y matemáticas a estudiantes de tercero y cuarto grados de primaria en distintos países de la región, incluyendo a México. Recientemente, en 2000, la OCDE, a través de su proyecto PISA, aplicó a alumnos de 15 años, esto es a los que concluyen la escuela secundaria, una batería de exámenes para analizar las habilidades y destrezas que se forman en la escuela en el campo de la lectura, matemáticas y ciencias. Sin embargo, fue sólo hasta el año 2001 cuando el gobierno de la República autorizó a tales organismos a hacer públicos los resultados.¹⁵

Aun así, podemos afirmar que los informes de evaluación son desconocidos por los docentes, por un número importante de especialistas y por el público en general. En México, a diferencia de otros países, no existe la estrategia de organizar talleres con los profesores para analizar el significado de tales reportes. Con ello, es prácticamente nulo el impacto que puedan tener en el mejoramiento de la educación.¹⁶

A lo anterior, hay que añadir la dificultad que tiene la ausencia de una teorización sobre la evaluación en México. La preocupación que orientó el establecimiento de las políticas al respecto se caracterizó por la definición de sus técnicas, indicadores y criterios de operación, pero no se fomentó la necesaria reflexión conceptual sobre los mismos. Por esta razón, las prácticas de evaluación no constituyen un sistema real de retroalimentación del sistema y del trabajo en el aula. La aplicación de exámenes masivos no va acompañada de una perspectiva de la dimensión formativa, sino que ha privilegiado sus aspectos técnicos y, en ocasiones, su sentido punitivo.

Mucho se ha hablado en este periodo de generar una cultura de la evaluación –más allá de lo incómodo de esta denominación para cualquiera que respete el pensamiento clásico de las humanidades, varios humanistas brincarían del susto al ver los usos que la era de la “modernización” ha dado al término “cultura”– debemos reconocer que la disciplina de la evaluación se realizó en diversos países, en general en los desarrollados, vinculada con un pensamiento eficientista, pero también ante la presión de identificar las mejores estrategias para optimizar la calidad educativa. Las prácticas de evaluación son abruptamente transplantadas o implantadas en nuestro medio educativo. La aplicación de la nueva generación de la evaluación se ha dado en prácticas educativas donde ha quedado ausente la dimensión teórica, en este sentido la llamada cultura de la evaluación en México ha sido bastante inculta.

Primer eje: la conceptualización del contenido en el EXANI-I y en la prueba PISA. Lectura y matemáticas

En este inciso analizaremos los fundamentos que subyacen en el manejo de dos áreas del contenido que son exploradas por ambas pruebas en dos tipos de áreas: español (lectura en el caso de PISA) y matemáticas. Para realizar esta tarea tomamos como referencia la información pública que sobre ambos test existe.

La conceptualización de contenido en el EXANI-I

En el caso del EXANI-I, podemos reconocer que se encuentra muy poca información, el mayor énfasis está dado sobre algunos reactivos. Sin embargo, los fundamentos de contenido del examen quedan escasamente perfilados de la siguiente manera, una prueba centrada en dos ejes: la exploración de habilidades intelectuales básicas y de conocimientos disciplinarios; las básicas son dos: verbal, donde se explora sin mayor argumentación –la comprensión de lectura, los sinónimos, antónimos y analogías–, y la habilidad matemática analiza el dominio de sucesiones numéricas, series espaciales, imaginación espacial y problemas de razonamiento. Por su parte, los conocimientos disciplinarios exploran específicamente conocimientos de español y matemáticas, además de los de otros campos del saber (cuadro 3).

La referencia para determinar el contenido del examen es el plan de estudios de primaria y secundaria interpretado por los especialistas, en

general profesores del bachillerato, lo que lleva a reproducir una perspectiva académica de los contenidos.

Llama la atención de manera significativa que no se den a conocer los supuestos conceptuales que llevaron a la decisión de la integración de un examen con estos elementos. Pero los mismos documentos a los que tuvimos acceso para este trabajo nos permiten observar que no existe una mayor conceptualización con respecto de las habilidades y de los conocimientos disciplinarios que son objeto de la examinación.¹⁷ A primera luz aparece que el elemento más significativo en la decisión para la integración de los 128 reactivos fue lograr una dosificación racional de 25% en la sección de habilidades y 75% en la de conocimientos.

CUADRO 3

Estructura del EXANI I

Área	Reactivos por área	%	Secciones	Reactivos por sección
HABILIDADES INTELLECTUALES BÁSICAS	32	25	Habilidad verbal	16
			Comprensión de lectura	7
			Sinónimos	3
			Antónimos	3
			Analogías	3
			Habilidad matemática	16
			Sucesiones numéricas	4
			Series espaciales	4
			Imaginación espacial	4
			Problemas de razonamiento	4
CONOCIMIENTOS DISCIPLINARIOS	96	75	Español	12
			Historia	12
			Geografía	12
			Formación cívica y ética	12
			Matemáticas	12
			Física	12
			Química	12
			Biología	12
Total de reactivos	128	100		128

Tomado de CENEVAL (2004a).

En habilidad verbal se establecen dos áreas: comprensión de lectura y manejo de vocabulario. A su vez, la primera está integrada por 10 habilidades: reconocer información, inferir hechos, identificar si un resumen contiene ideas principales, completar un cuadro sinóptico, identificar conclusión, identificar secuencias de acontecimientos, establecer relaciones causa, diferenciar hechos y opiniones, etcétera. Por su parte el manejo de vocabulario se centra en tres aspectos: establecer analogías, distinguir palabras y expresiones con significado opuesto y similar (CENEVAL, 2004a:90).

Una descripción un poco más amplia se encuentra en otro documento del Centro, donde se establece que:

La comprensión de textos es una actividad mental que requiere de concentración total para su realización eficiente. Comprende el reconocimiento del significado de los vocablos, el descubrimiento de las relaciones entre las frases y las oraciones, así como de los distintos párrafos que integran el texto. Considera también la identificación de las ideas fundamentales y su captación para emplearlas posteriormente conforme a una necesidad determinada como, por ejemplo, decir de qué trata un texto o a qué conclusión conduce (CENEVAL:2004b:37).

La *Guía del examen* que se proporciona electrónicamente establece que se trata de hacer una exploración amplia de conocimientos que se requieren para un mejor desempeño en el siguiente nivel educativo. Al término de la misma se presenta un temario para orientar el estudio que, de alguna forma, constituye la respuesta más cercana a la opción de contenido que se asume.

En el temario se mencionan cuatro habilidades matemáticas: sucesiones numéricas, series espaciales, imaginación espacial y problemas de razonamiento. El temario de matemáticas es mucho más clásico: *a)* aritmética (números naturales y enteros, fracciones decimales, proporcionalidad); *b)* álgebra (monomios y polinomios, ecuaciones, plano cartesiano y funciones); y *c)* geometría (ángulos entre paralelas y una secante, triángulos, semejanza, etcétera).

De esta manera podemos afirmar que esta información muestra una débil concepción del contenido, tendiendo a privilegiar la estructura académica que se observa en los planes de estudio.

La conceptualización de conocimientos y habilidades en lectura y matemáticas en la prueba PISA

La prueba PISA dedica un amplio apartado a definir lo que se entiende por lectura, en el instrumento aplicado en 2000, y por matemáticas, en el de 2003.

Un análisis del reporte 2000 que emite PISA sobre sus exámenes permite identificar que, de ocho capítulos, dos están destinados a la conceptualización del contenido y uno se dedica a la denominada aptitud para la lectura. Los autores de la prueba establecen, con claridad, tres funciones que se pueden identificar en las actividades lectoras: obtención de información, interpretación de textos, y reflexión y evaluación de la información. A partir de ellas establecen lo que consideran cinco niveles de dificultad, de suerte que en la elaboración de los reactivos del examen y en el reporte de evaluación desarrollan preguntas para cada función y nivel de complejidad del instrumento de evaluación (cuadro 4).

CUADRO 4

Niveles de dificultad de las escalas

NDC	Funciones		
	<i>Obtención de información</i>	<i>Interpretación de textos</i>	<i>Reflexión y evaluación</i>
5	Ubicar, ordenar y combinar fragmentos anidados profundamente y que se encuentran fuera del texto. Tratar con información extensa.	Interpretar significado del lenguaje y demostrar una comprensión total y detallada.	Emitir juicios e hipótesis , apoyándonos en un conocimiento especializado y en el entendimiento de textos largos.
4	Ubicar y ordenar fragmentos de acuerdo con varios criterios y con contextos y formas no familiares.	Interpretar significados de una sección en relación a todo el texto.	Emplear conocimientos formales o públicos en la comprensión precisa de un texto.
3	Ubicar y reconocer la relación entre fragmentos.	Identificar, una idea central y significado de frases o palabras.	Realizar conexiones, dar explicaciones con base en características.
2	Ubicar fragmentos de información.	Inferencias de bajo nivel.	Realizar comparaciones en relación al conocimiento exterior.
1	Tomar en cuenta un solo criterio para ubicar fragmentos independientes expresados explícitamente.	Reconocer tema central o propósito del autor en un tema familiar.	Realizar conexiones simples entre información y conocimiento común.

Adaptado de OCDE, 2001:38.

A este planteamiento se añaden otras precisiones, determinan que los reactivos del examen buscan estudiar la habilidad desarrollada en el trabajo sobre textos continuos (un libro, un artículo) y textos discontinuos (instrucciones para el armado de algún objeto). De la misma manera se reconoce que los textos tienen diversos usos por su conformación: uno privado (una novela), uno público (un anuncio), otro ocupacional (un informe de trabajo) y un uso escolar (un documento de estudio).

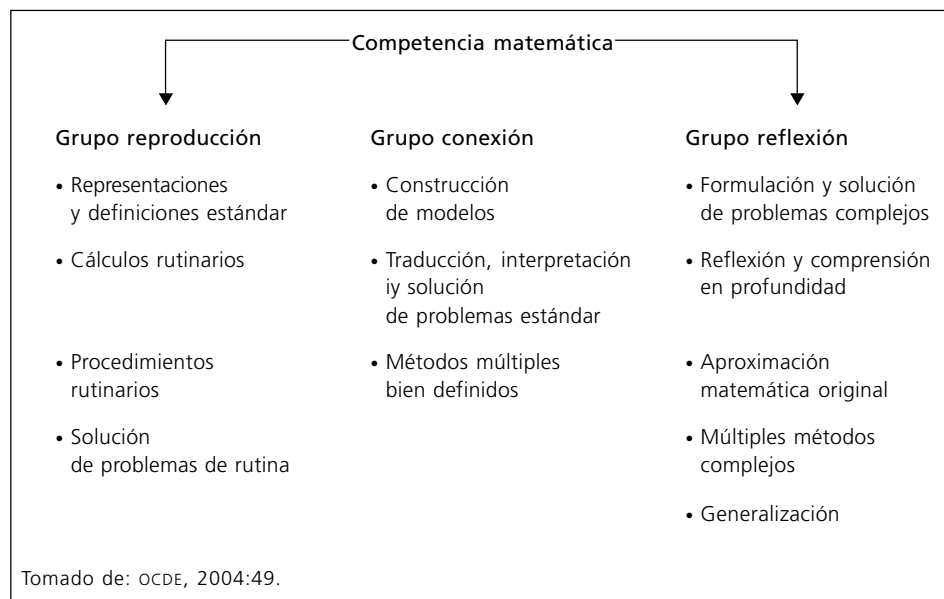
Se puede afirmar que en la propuesta de PISA, las habilidades lectoras son complejas. De suerte que para reconocer si un estudiante ha desarrollado una habilidad lectora requiere realizar alguna precisión sobre si el mismo se refiere a un resultado promedio de todas las funciones y niveles de complejidad que fueron objeto de la prueba o si, por el contrario, sólo se refieren a uno de ellos.

Esto significa que los resultados que se publicitan del examen son mucho más complejos de lo que habitualmente se suele manejar: afirmar que los jóvenes mexicanos obtuvieron el lugar 32 en la lectura, en el futuro no significa nada. El mismo informe previene sobre ello y plantea una información mucho más desagregada de las habilidades que son objeto de indagación.

PISA dedicó la prueba 2003 a explorar con mayor profundidad las destrezas desarrolladas en el campo de las matemáticas. En este caso, la OCDE decidió realizar diversas publicaciones, una de ellas se titula *Marcos teóricos de PISA 2003. Conocimientos y destrezas matemáticas, lectura, ciencias y solución de problemas*. En este documento se dedica un capítulo a cada uno de estos aspectos. Establece que el área de las matemáticas “se ocupa de la capacidad de los estudiantes para analizar, razonar y comunicar ideas de modo efectivo, al plantear, formular, resolver e interpretar problemas matemáticos en diferentes situaciones” (OCDE, 2004:27).

La construcción del examen considera cuatro ángulos sobre los cuales se construyeron las preguntas con distinto nivel de dificultad. Éstos son: *a)* el reconocimiento de que las habilidades matemáticas se desempeñan en cuatro situaciones diversas (personales, educativas y profesionales, públicas y científicas); *b)* el establecimiento de tres tipos de competencia matemática (de reproducción –aplicar reglas–, de conexión y de reflexión); *c)* el manejo de cuatro ideas principales (cantidad, espacio y forma, cambio, relaciones e incertidumbre); y *d)* finalmente en este contexto la definición de que el examen contendrá en partes iguales un tercio de preguntas de opción múltiple, de respuesta cerrada y de respuesta abierta (cuadro 5).

CUADRO 5

Representación sintética de los grupos de competencia matemática

A ello hemos denominado, en nuestra aproximación, una concepción de contenido para la elaboración del examen. Tradicionalmente en las pruebas estandarizadas se emplea el término “validez del contenido” la que, en general, se determina a partir de un juicio de expertos. El concepto estadístico que subyace a la tradicional denominación de validez del contenido resulta completamente insuficiente cuando se vislumbran las perspectivas que enunciamos.

Ésta es la primera diferencia significativa que encontramos entre los exámenes mexicanos y los internacionales, ello nos ha llevado a afirmar que en estos últimos existe una teoría del contenido que no se vislumbra en los nacionales.

Segundo eje:**el concepto de aprendizaje que subyace en ambas pruebas**

Existen varias formas para analizar el concepto de aprendizaje que subyace en ambas pruebas. En este punto también encontramos que el planteamiento de PISA es más explícito y, como veremos más adelante, se refiere

a un desarrollo personal a partir de los conocimientos que se aprenden en la escuela. Esta visión del aprendizaje seguramente guarda relación con la forma como se construyeron ambos exámenes.

El EXANI-I es una prueba construida a partir del criterio de los expertos. Un buen número de ellos son profesores de su especialidad que laboran en el sistema educativo, esto garantiza un buen dominio de la información con la que construyen cada pregunta del examen, pero también deja vislumbrar la visión de aprendizaje cercana a la retención y uso de la información que, en general, priva en el sistema educativo mexicano. El otro referente para la construcción del examen son los contenidos de las diversas materias que integran el plan de estudios de la escuela secundaria.

Por su parte, la prueba PISA refleja otra dinámica. En primer término, no existe un plan de estudios de referencia, pues la diversidad de países en donde se aplica el examen impide que ése sea el criterio que sirva de partida para su elaboración. En segundo término, su construcción refleja varios elementos del debate de las reformas educativas que se están impulsando en Europa (EURYDICE, 1997), de la manera como se está modificando el concepto de aprendizaje escolar a través de la discusión e incorporación de una noción del mismo basada en competencias (EURYDICE, 2002). Tema que aleja el aprendizaje del empleo de la función de recuerdo y uso aplicado de la información. Por el contrario, a partir de un cuestionamiento al ejercicio rutinario escolar, la perspectiva de las competencias abre un escenario distinto para el trabajo escolar.

Reactivos de habilidades verbales, español y comprensión de lectura del EXANI-I

Una manera como se puede analizar la forma en que este concepto de contenido se materializa en los exámenes es revisando alguno de sus reactivos. No obstante, también aquí surge un problema, mientras PISA ha publicado 300 reactivos de su banco, lo que significa que quedan congelados para su uso en las pruebas, el CENEVAL publica un número muy reducido de los mismos.

Los tipos de reactivos que se emplean en cada examen no dejan lugar a dudas sobre una visión muy mecánica de la lectura: identificar información, así en la medición de habilidades verbales del EXANI-I, se encuentra el reactivo siguiente:

PREGUNTAS EXANI-I

Hace 2,500 años, sobre estas mismas nieves invernales, y sobre sus sucesoras flores amarillas de primavera, cabalgaba un pueblo preparado para la guerra, la caza y el pastoreo como ningún otro en la región: los pazyryk, una estirpe de valientes jinetes que vivían en la estepa. Este clan estaba emparentado con los escitas, antiguo pueblo seminómada que provenía de las regiones asiáticas donde actualmente se encuentra Irán.

Hoy, los miembros de un equipo científico del Instituto Ruso de Arqueología y Etnografía de Novosibirsk desentierro uno a uno, fascinador y con mano trémula, los vestigios más sorprendentes de aquella civilización. Entre ellos, las dos joyas más valiosas son las momias de una dama de la alta sociedad pazyryk (siglo VI antes de Cristo) y de un joven jinete enterrado junto a su caballo (siglo III o II antes de Cristo).

Este último, que es el hallazgo más reciente, permite a los expertos albergar esperanzas sobre la posibilidad de que haya más restos momificados escondidos en las frías tierras de Ukok.

La momia del joven escita fue encontrada con sus vestimentas de caballero –gorro, abrigo de marmota, pantalones rojos y botas de fieltro– casi intactas. Su cabeza estaba adornada por dos largas trenzas pelirrojas que, sorprendentemente, han desafiado incólumes el paso de más de dos mil años bajo tierra, y en la piel de la espalda y el pecho tenía tatuados dibujos de un gran alce.

Este último dato contradice algunas teorías históricas sobre el papel de los tatuajes en las antiguas comunidades euroasiáticas. Tradicionalmente se había creído que los escitas y otros pueblos utilizaban los adornos en la piel como símbolo de nobleza o posición jerárquica elevada. Sin embargo, este joven recientemente desenterrado era un simple guerrero nómada sin abolengo. ¿Por qué estaba también tatuado? El estudio de posibles futuros restos podrán, quizás, dar una respuesta a esta pregunta.

Mientras tanto, su momia ya ha sido trasladada a Moscú, al mismo laboratorio en donde se conserva el cuerpo embalsamado de Lenin, para que sus técnicos estudien cómo mantener los restos de este hombre fuerte y alto que tenía 30 años de edad cuando murió.

1. De la región de Irán provienen los:

- a) Escitas
- b) Rusos
- c) Nómadas
- d) Pazyryk
- e) Euroasiáticos

2. Las características más sobresalientes de la momia son:

- a) El cabello y el tatuaje
- b) La vestimenta y la piel
- c) La vestimenta y las joyas
- d) El gorro y las botas
- e) El abrigo y los pantalones rojos

Tomado de CENEVAL (2004a).

También podríamos tomar algunos ejemplos de los reactivos que se emplean en la medición de conocimientos de español, como el del caso siguiente:

PREGUNTAS EXANI-I

1. En el siguiente párrafo se relata un hecho que ya ocurrió pero hay problemas de correlación verbal. Señala la opción que indica las correlaciones correctas

La banda **dejó** de tocar cuando apenas **empieza** a llover. Los músicos **habían entrado** a la casa principal donde se les **ofrece** un almuerzo. **Hubo** más comensales de los esperados y la comida no **alcanzaría** para todos.

- a) Dejaría/empezaría/entrarían/ofrecería/había/alcanzaría
- b) Dejó/empezaba/entraron/ ofreció/hubo/alcanza
- c) Dejaba/empezaba/entraban/ofrece/había/alcanzaba
- d) Había dejado/había empezado/habían entrado/ había ofrecido/había/había alcanzado
- e) Deja/empieza/entran/ofrece/hay/alcanza

2. ¿Cuál es la categoría gramatical de las palabras escritas en negritas en el siguiente fragmento?

La **contemplación** amorosa y devota de las **maravillas** del suelo nos dejó alguna vez **impresiones** encantadoras y gratas, que conservamos con cariño durante **nuestra vida**.

- a) Sustantivos
- b) Adjetivos
- c) Adverbios
- d) Conjunciones
- e) Preposiciones

Tomado de CENEVAL (2004a).

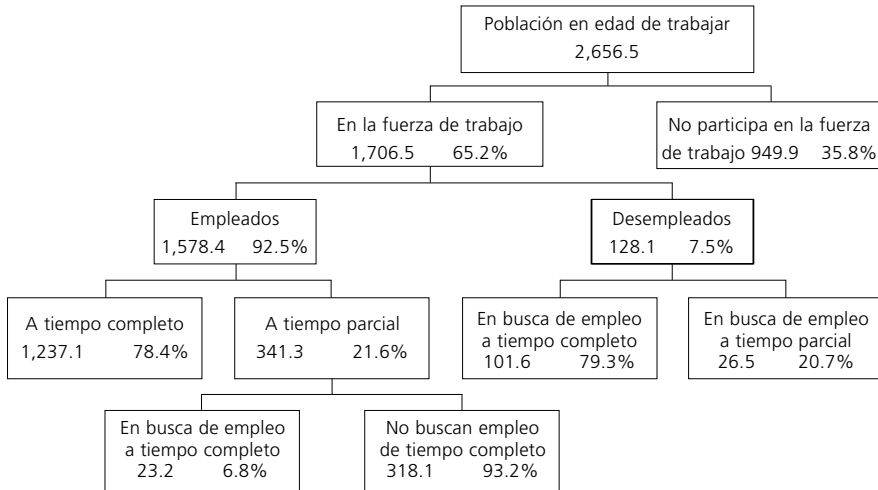
Es importante preguntarse sobre el tipo de aprendizaje que subyace en ambas preguntas, en particular, es necesario resaltar las habilidades intelectuales que demandan del estudiante.

Reactivos de habilidad lectora del PISA

En el caso de la prueba del PISA, las preguntas tipo que se aplican en este instrumento parten de otra perspectiva. Lo podemos mostrar con algunos ejemplos:

PRUEBA PISA

El diagrama que ve abajo muestra la estructura de trabajo de la fuerza productiva de un país o la población en edad de trabajar. La población total, en 1995 era de 3.4 millones de personas



PRUEBA PISA

Pregunta

¿Cuántas personas en edad de trabajar no se encontraban en la fuerza de trabajo? (Escribe el No. de personas no el porcentaje)

Escala

Obtención de información

N

4 y 3

¿Cuáles son los dos principales grupos en que se divide la población en edad de trabajar?

Interpretación de textos

Menos de 1

- a) Empleados y desempleados
- b) En edad de trabajar y no en edad de trabajar
- c) Empleados a tiempo completo y a tiempo parcial
- d) En la fuerza de trabajo y fuerza de la fuerza de trabajo

El diagrama jerárquico se eligió para presentar la información sobre la estructura de la fuerza de trabajo, probablemente porque es particularmente útil para:

Reflexión y evaluación

3 y 2

- a) Los cambios en el tiempo
- b) El tamaño de la población total
- c) Categorías dentro de un grupo
- d) El tamaño de cada grupo

Tomado de: OCDE, 2001.

Reactivos de habilidades matemáticas del EXANI-I

Menos evidente son los supuestos de aprendizaje que subyacen en cada una de las preguntas de matemáticas. Sin embargo, es diverso el grado de complejidad de los cuestionamientos de ambas pruebas.

HABILIDADES Y CONOCIMIENTOS EN MATEMÁTICAS. EXANI-I

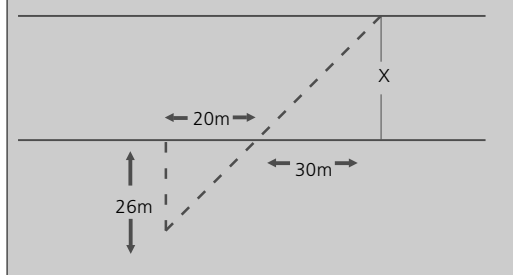
1. Observa la siguiente sucesión: 4, 9, 14, 19, 24 ... ¿Qué número ocupará el lugar 34 de la sucesión?
 A) 39
 B) 164
 C) 169
 D) 170
 E) 174
2. Selecciona la opción que contenga el término que sigue en la sucesión presentada.
 2, 4, 9, 20 _____
 A) 39
 B) 40
 C) 41
 D) 42
 E) 43
3. Un niño tiene el mismo número de hermanas que de hermanos, y una de sus hermanas tiene la mitad de hermanas que de hermanos. ¿Cuántos niños hay en la familia? ¿Cuántos son hombres y cuántas mujeres?
 A) 5, 3 hombres y 2 mujeres
 B) 4, 2 hombres y 2 mujeres
 C) 5, 2 hombres y 3 mujeres
 D) 7, 4 hombres y 3 mujeres
 E) 7, 3 hombres y 4 mujeres
4. Soluciona la siguiente ecuación $3x - 1 = x + 3$
 A) $x = 0.5$
 B) $x = \frac{4}{5}$
 C) $x = \frac{4}{7}$
 D) $x = 2$
 E) $x = 5$
5. Para hornear un pavo se considera que por cada $\frac{1}{2}$ kg se requieren $\frac{3}{4}$ de hora a fuego, ¿durante cuánto tiempo se debe hornear un pavo de cinco kg?
 A) 6 horas 45 min.
 B) 7 horas
 C) 6 horas 30 min.
 D) 7 horas 15 min.
 E) 7 horas 30 min.

(CONTINÚA)

6. Una empresa de investigación privada aplicó una encuesta a 200 familias, de las cuales 32 dijeron tener un hijo; 55 dos hijos, 58 tres hijos, 25 cuatro hijos y 30 cinco o más hijos ¿Cuál es la probabilidad de que una familia escogida de la encuesta al azar tenga a lo más tres hijos?

- A) 71.5%
- B) 56.5%
- C) 27.5%
- D) 45%
- E) 43.5%

7. ¿Cuál es el ancho del río?



- A) 45 metros
- B) 17.3 metros
- C) 39 metros
- D) 23 metros
- E) 36 metros

Tomado de: CENEVAL (2004a).

En estas preguntas se pueden observar dos aspectos: se trata de interrogantes que se suelen emplear en el sistema de enseñanza y, en este caso, muestran que el conocimiento matemático puede emplearse para resolver problemas que se encuentran en el medio social donde se desenvuelve el estudiante. Sin embargo, la perspectiva de estas preguntas es eminentemente académica.

El examen cumple con establecer preguntas con base en un temario que ofrecido al estudiante. Las preguntas 1, 2, 4 y 7 corresponden a la aplicación de conocimientos matemáticos, mientras que las otras requieren que el estudiante pueda traducir, para su resolución, conceptos del lenguaje común a términos algebraicos. El examen no ofrece otra interpretación particular sobre las mismas.

Reactivos de habilidades matemáticas en PISA

En el caso de la prueba PISA encontramos diversas pistas para analizar las preguntas que son la base de su examen.

EXAMEN PISA

Se ingresan 1,000 zeds en una cuenta de ahorro en un banco. Existen dos opciones: o bien obtener un interés anual de 4%, o bien obtener una prima inmediata de 10 zeds y un interés anual del 30%. ¿Qué opción es mejor al cabo de un año? ¿Y al cabo de dos años?

Tomado de: OCDE (2004).

En el reporte del examen no queda claro si la moneda objeto de esta prueba se adapta a la que existe en cada país o si se emplea esa denominación genéricamente, en cuyo caso sería uno de los factores que añaden un ingrediente de dificultad al mismo examen. No es este el espacio para discutir tal cuestión, pero esa dificultad también aparece en español, cuando seleccionan como objeto de la lectura una cámara fotográfica profesional con diversas especificaciones en su lente; teniendo en cuenta que la mayoría de los estudiantes realmente no tienen acceso cotidiano a esta situación. En el fondo, PISA no informa si cumple con la obligación técnica de tipificar la prueba para cada idioma y país.

Otro tipo de reactivos guardan relación con la clasificación de las competencias matemáticas que han definido: reproducción, conexión y reflexión. Los casos serían los que se muestran en la tabla de la página siguiente.

Sólo deseamos enfatizar la forma como una prueba, el EXANI-I, recurre a la manera habitual en que los docentes trabajan las matemáticas en el salón de clase, a los dos tipos de preguntas que formula: unas de reproducción de información y otras de desarrollo de pensamiento algebraico. En varios puntos coincide esta prueba con la elaborada para el PISA, aunque su principal diferencia se encuentra en el concepto de aprendizaje que orienta toda la prueba PISA.

Mientras el EXANI-I toma como referencia lo que debe ser aprendido en la educación básica y lo que permitirá al estudiante realizar sus estudios de bachillerato. Buscando que los reactivos se asemejen lo más posible a los reactivos escolarizados, la prueba PISA parte de un presupuesto totalmente diferente: analizar el desarrollo de habilidades y destrezas para la vida. En una parte de su documento hacen referencia a que no bastan los conocimientos adquiridos en la escuela para salir exitoso en el examen, pues además

de esos conocimientos se requiere haber desarrollado una serie de habilidades y destrezas para utilizar esos conocimientos en la vida cotidiana. En esto son enfáticos, sin conocimientos no se resuelve el examen, pero se requiere de haber desarrollado otras habilidades.

EXAMEN PISA

Preguntas del grupo de reproducción

1. Resuelve la ecuación
 $7x - 3 = 13x + 15$

2. ¿Cuál es la media de 7, 13, 8, 14, 15, 97?

Preguntas del grupo de conexión

3. María vive a dos kilómetros de su colegio y Martín a cinco. ¿A qué distancia viven el uno del otro?

4. Los dos siguientes anuncios aparecieron en un diario de un país, cuya unidad monetaria es el zed.

EDIFICIO A

Se alquilan oficinas
 58.95 metros cuadrados
 479 zeds al mes
 100-120 metros cuadrados
 800 zeds al mes

EDIFICIO B

Se alquilan oficinas
 35-260 metros cuadrados
 90 zeds por metro cuadrado al año

Si una empresa está interesada en alquilar una oficina de 110 metros cuadrados en ese país durante un año, ¿en qué edificio de oficinas, A o B, deberá alquilar la oficina para conseguir el precio más bajo? Escribe tus cálculos.

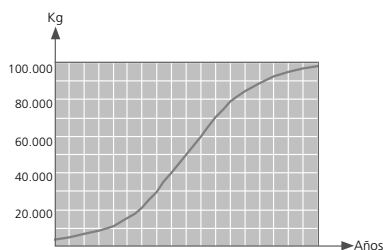
[IEA/TIMSS]

Preguntas del grupo de reflexión

5. En un determinado país, el presupuesto nacional de defensa fue de 30 millones (en la moneda del país) en 1980. El presupuesto total de este año fue de 500 millones. Al año siguiente, el presupuesto de defensa pasó a 35 millones, mientras que el presupuesto total fue de 605 millones. La inflación del periodo comprendido entre los dos presupuestos alcanzó el 10 por ciento.

- a) Te invitan a dar una conferencia en una asociación pacifista. Intentas explicar que el presupuesto de defensa ha disminuido en este periodo. Explica cómo lo harías.
 b) Te invitan a dar una conferencia en una academia militar. Intentas explicar que el presupuesto de defensa ha aumentado en este periodo. Explica cómo lo harías.

6. Se repobló con peces un canal fluvial. El gráfico muestra un modelo de cómo ha crecido el peso total de los peces en el canal fluvial.



Tomado de: OCDE (2004).

Podemos suponer que cuando los estudiantes mexicanos se enfrentan a una y otra prueba, se enfrentan a una problemática distinta. Ello podría ser uno de los factores que explique el bajo rendimiento que obtienen en la prueba internacional.

No podemos cerrar este rubro sin reconocer que mientras los reactivos de las pruebas mexicanas se manejan con mucha discreción, el proyecto PISA ha publicado más de 300 reactivos de los que componen su prueba (OCDE, 2002). El tema es bastante claro, lo que se busca es que se disponga de un elemento que permita contribuir al análisis de los mismos.

Tercer eje: la concepción de evaluación y teoría del test en la que se apoya cada prueba

El tercer elemento de nuestro análisis apunta a señalar la concepción de evaluación que subyace en cada una de estas pruebas. En un primer sentido busca una respuesta sobre la finalidad de la evaluación. Atender el para qué se evalúa para, posteriormente, trabajar algunos de los fundamentos con los cuales se realiza la tarea de la evaluación.

De alguna forma hemos mencionado que el objetivo de la evaluación es diverso en cada una de estas pruebas. El EXANI-I establece que con esta actividad se busca diferenciar de manera confiable y precisa el nivel académico de los sustentantes. Las preguntas corresponden una amplia gama de contenidos, “evaluar lo que se sabe y cómo se usa lo que se sabe, es decir las habilidades” (CENEVAL, 2004a:4-11), que deben haber desarrollado durante sus estudios. Finalmente se concluye que examen se califica dando un punto a cada reactivo cuya respuesta es correcta. “El CENEVAL únicamente califica y emite reportes de resultados correspondientes a las instituciones y usuarias del examen” (CENEVAL, 2004a:9). En todo caso son las instituciones las que toman diversas decisiones con respecto a la información recibida y quienes, en general, sólo la emplean para decidir sobre quiénes pueden ser admitidos como estudiantes.

En estricto sentido no hay ninguna declaración sobre la teoría del test que orienta la construcción de estos exámenes. Algunas afirmaciones llevarían a suponer que la visión sigue centrada en la teoría clásica del test, sobre todo por la perspectiva que tienen del manejo de la distribución de frecuencias respecto del número de aspirantes. En la *Guía* para preparar el examen afirma: “incluye preguntas cuya gama de dificultad va desde las muy fáciles (aquellas que la gran mayoría de los sustentantes resuelve co-

rectamente) hasta las muy difíciles (aquellas que sólo los sustentantes muy capaces pueden responder correctamente). Es previsible que la mayoría de los sustentantes alcancen puntajes cercanos al 50%” (CENEVAL, 2004a:5).

Desde la perspectiva de la evaluación se puede afirmar que el CENEVAL no asume de una manera amplia la función de retroalimentación de la evaluación. Ciertamente en los últimos años se ha orientado más hacia esta tarea, pero difícilmente un profesor de secundaria de matemáticas, español o ciencias tendrá elementos para mejorar su trabajo docente como resultado de la información de este examen. Se suele argüir que el examen no permite esta información porque ha sido concebido en otra perspectiva, sin embargo, la que brinda puede ir mucho más de lo que actualmente se ofrece. Identificar las secundarias donde los estudiantes obtienen mejores resultados innegablemente que es una tarea importante, pero también podrían informar sobre los temas en los que tienen mayores dificultades para su desempeño. Si un temario permite orientar la construcción del examen, éste podría ser el punto de referencia para una interpretación de resultados que apunte hacia la manera como se desempeña una población escolar en un momento determinado.

Los informes del CENEVAL están centrados en dos aspectos: la determinación de la posición global y particular que tiene una institución o un estudiante de acuerdo con un desempeño general. De esta manera se manifiesta el lugar en el que se encuentra un atributo “habilidad verbal” con relación a una media. Ello también permite al Centro informar sobre la posición que tiene la institución educativa respecto de las otras. Con independencia del uso que las instituciones den a esta información, a la fecha el Centro no ha podido elaborar un informe que contenga elementos pedagógicos directamente orientados al docente. Por su condición, varios de ellos tienen un carácter restringido.

Por su parte, la prueba PISA tampoco hace una serie de declaraciones específicas sobre el manejo técnico de la evaluación. Lo que obliga también a realizar una inferencia del concepto de evaluación y la teoría del test en los que se apoya. En este aspecto encontramos una diferencia significativa, existe mayor desarrollo en los reportes que se presentan como resultados de su aplicación. Los reportes tienen un carácter público, si bien son poco conocidos y estudiados por los especialistas, por los maestros y por los padres de familia en nuestro medio, con lo cual, tampoco tienen un impacto en la mejora del trabajo en el aula.

En *Conocimientos y aptitudes para la vida* (OCDE, 2001) se establece que la prueba PISA “evalúa los niveles de una amplia gama de conocimientos y aptitudes alcanzados por jóvenes de 15 años de los principales países industrializados” (OCDE, 2001:17). Para la evaluación de cada una de las áreas de aptitud se elaboró un marco de referencia que examina esta prueba. Con el fin de poder realizar la comparación se buscó que hubiera semejanza en la población objetivo: estudiantes de 15 años inscritos en el sistema escolar. El informe que se elaboró presenta:

[...] distribuciones de desempeño para cada país y no sólo notas promedio, se emplea información de apoyo sobre los estudiantes, sus escuelas y sus sistemas de educación para analizar toda una serie de factores asociados con los distintos niveles de desempeño.... con ello se busca que PISA sea una herramienta poderosa para mejorar la comprensión de lo que fomenta el éxito en la educación (OCDE, 2001:17).

En el informe de 2004 también se formula la precisión de que la prueba tiene preguntas de diferentes tipos: unas requieren que los alumnos “seleccionen o elaboren respuestas simples que pueden ser comparadas directamente con una respuesta única como preguntas de respuesta múltiple o preguntas de respuesta cerrada”. El documento considera que estas preguntas no sólo tienen una respuesta cerrada, sino que evalúan destrezas de orden inferior. “Otras preguntas más constructivas, requieren que los alumnos elaboren su propia respuesta y han sido ideadas para poder medir aspectos más generales que los que suelen recoger otros estudios más tradicionales” (OCDE, 2004:22). Además, se aplica un cuestionario sobre el entorno de los alumnos y de los directores de los centros educativos.

Todo ello nos lleva a afirmar que existe una perspectiva de evaluación mucho más amplia en la prueba PISA. Esta amplitud se refleja tanto en el concepto global de la evaluación y su sentido formativo, como en la concepción del instrumento de evaluación. Los reportes que emite la OCDE, no dejan dudas al respecto. Se trata de una información amplia y al mismo tiempo puntual. El marco teórico en el que es definido el contenido de cada una de las áreas de evaluación se constituye en el referente para analizar los resultados. En vez de una puntuación global se entrega un reporte internacional, con datos nacionales sobre los desempeños de los estudiantes en cada una de las tres áreas de conocimiento exploradas y en cada uno

de los niveles de conocimiento. Al mismo tiempo el reporte ofrece una interpretación contextual sobre las condiciones que favorecen el aprendizaje de los estudiantes.¹⁸

Por otra parte, aunque tampoco es explícito en los documentos de la OCDE, se puede inferir que la parte de los exámenes conformada con reactivos de opción múltiple se apoya en la teoría de respuesta al ítem (Muñiz, 1997), con elementos de desempeño criterial. Ello resulta, de alguna manera, en los reportes que ha presentado donde la referencia es al dominio que muestran los estudiantes en cada uno de los niveles o ámbitos de conocimiento, en los que se ha definido el área de conocimiento que es objeto de exploración. La comparación se realiza, entonces, sobre la forma como estos niveles se establecen internacionalmente en los distintos países.

La perspectiva formativa se encuentra presente en las diversas tareas asociadas con la evaluación: la presentación de un informe público y la publicación de parte del banco de reactivos con la finalidad de ofrecer una evidencia puntual sobre la forma en la que cada concepto es manejado.

Conclusiones

La OCDE, a través de Programa Internacional de Evaluación de Estudiantes, ha realizado una significativa aportación a la teoría y práctica de la evaluación. En primer término, mostrando que el papel que tiene una teoría del contenido en la elaboración de los exámenes es que puede enriquecer las perspectivas que se suelen derivar únicamente de la teoría del test. La teoría del contenido, a su vez, se enriquece con la noción de aprendizaje que se asuma en la construcción de un instrumento de evaluación; de esta manera, el problema de la elaboración de pruebas no se puede remitir sólo a una dimensión técnica.

De igual manera, la prueba PISA ofrece un panorama de la evaluación más amplio, que fundamentalmente recupera la dimensión formativa de la tarea de evaluar. No basta con asignar un puntaje a un examen, cuando lo que se requiere es tomar otras decisiones en el ámbito pedagógico.

De alguna forma será necesario volver a tener presente que la finalidad de la evaluación no se encuentra en sí misma, ni circunscrita sólo a dar resultados. Por demás es preocupante que las instancias responsables de la tarea de la evaluación se preocupen de generar talleres para que los medios de comunicación entiendan la evaluación y no se preocupen por hacer

llegar su información a los docentes de las instituciones educativas. La mejora de la calidad de la enseñanza, uno de los objetivos implícitos en la evaluación, necesariamente supone la modificación de las prácticas cotidianas de docentes y alumnos. En este cambio, que por sus mismas circunstancias puede quedar atrapado en la rutina, es precisamente donde la tarea de la evaluación debe proporcionar información que permita revisar, entender y establecer nuevas alternativas a estas prácticas. En cierto sentido es necesario invertir la ponderación de los destinatarios, colocando a los docentes y alumnos en el centro de la preocupación por el conocimiento de sus resultados. La sociedad, las autoridades educativas evidentemente que son destinatarios de los resultados de la evaluación, pero teniendo presente que el centro de todo proyecto de reforma pasa necesariamente por el aula.

En este momento los reportes de la OCDE constituyen la mejor muestra de que es factible realizar una tarea formativa desde el trabajo de la evaluación. Ello puede ser un elemento que impulse a desarrollar este carácter en otras pruebas que se aplican a nuestros estudiantes.

Si bien no se pueden comparar los resultados de los estudiantes mexicanos en las dos pruebas que han sido objeto de análisis en este artículo, es claro que estudiantes de mismas condiciones de escolarización presentan exámenes que tienen presupuestos de construcción diversos. Ello explicaría, parcialmente, la diferencia que puede existir entre los resultados obtenidos en una prueba y en otra. Pero también invita a una reflexión que no es secundaria, referida a los contenidos y a los tipos de aprendizaje que se deben fomentar en el sistema educativo.

De alguna forma es una invitación para pensar las lógicas que acompañan las reformas educativas en el país. Lógicas de cada seis años, donde no se da el tiempo suficiente para que alguna idea orientadora sea realmente asumida por los docentes, lógicas de reformas educativas por decreto, ante innegables deficiencias de formación del magisterio.¹⁹ Si deseamos, como país, un mejor logro educativo será necesario nuevamente reconocer al aula como espacio de esta tarea.

Otra aportación significativa del manejo de la prueba de la OCDE es la forma como concibe su función de retroalimentación y su tarea formativa. Evidentemente, el informe es extenso por la cantidad de temas que ofrece y por el trabajo puntual sobre cada uno de los aspectos que forman parte del examen. Es el informe el que permite trabajar el concepto de contenido,

de aprendizaje y de evaluación del que parte el examen, lo cual indudablemente ayuda a construir los resultados, así como es un factor clave en la interpretación de los mismos. No pueden faltar en el mismo una apreciación sobre los factores contextuales del examen. Este aspecto, en general, es poco trabajado en las evaluaciones nacionales.

Pero es necesario tener presente que no basta con elaborar un informe, sino que es indispensable establecer un mecanismo que garantice que los docentes lo conozcan y reflexionen sobre el mismo. Hacer una afirmación genérica no ayuda a mejorar al sistema educativo. Cuando el manejo de la información se deja en manos de los medios de comunicación, se cumple con una premisa de la teoría de la reforma educativa impulsada por el gobierno de Reagan: hacer que la sociedad asuma el papel de impulso al cambio de la educación a partir del conocimiento de sus deficiencias (Cramer, 1983:7-10). Lo cual no necesariamente acontece en nuestro medio. Es imprescindible tener presente que el cambio en la educación sólo será posible en tanto haya mayor conciencia de los problemas de funcionamiento de la educación por parte de los docentes y cuenten con una mejor formación para enfrentar esas deficiencias. El objetivo final de la evaluación debe quedar en manos de los docentes. Ésta es una de las grandes agendas pendientes de las perspectivas de evaluación que se aplican en el país.

En otra línea de conclusiones, este trabajo muestra que cada prueba masiva tiene supuestos distintos sobre el contenido, el aprendizaje y la evaluación, incluyendo una visión distinta de la teoría del test. Este aspecto no es analizado en el caso nacional. Pareciera que los estudiantes se enfrentan a sistemas de medición que son idénticos, o que tienen parámetros intercambiables, lo cual no es cierto. Los resultados de una prueba son válidos en el marco de los supuestos que orientaron dicha prueba, de esta manera sería necesario interpretar los resultados de un examen en el marco del mismo.

Pero esta afirmación tiene otras consecuencias. Si se desea que los estudiantes mexicanos salgan exitosos en un examen será necesario prepararlos en aquellos conocimientos y habilidades que forman parte de la estrategia de medición de tal prueba. Este tema es particularmente relevante porque en la actualidad existen diferencias significativas entre los exámenes nacionales y las pruebas internacionales que se están aplicando en el sistema educativo.

Notas

¹ Claramente se puede observar cómo la literatura del campo de la evaluación de los años cincuenta y sesenta se refería fundamentalmente al aprendizaje, mientras que en la actualidad lo hace a la evaluación de las instituciones, de los programas y de los actores. De tal manera, este proceso se ha convertido en un campo de conocimiento complejo, los métodos y sistemas que se emplean para “medir el aprendizaje” —la teoría del test en particular— no tiene un impacto significativo en los demás ámbitos de la evaluación.

² En mi opinión ocultan otro problema también fundamental: la desconfianza que las políticas, las autoridades y un grupo de especialistas educativos tienen hacia el docente. En el fondo se considera que desde él no se puede generar un mejoramiento de la educación. Por ello, por ejemplo el Banco Mundial en su documento de política para la escuela primaria de 1990 ponía mayor énfasis en el gasto en materiales educativos que en los procesos de formación y remuneración económica de los docentes. Cfr. Banco Mundial (1992).

³ Es importante recordar que en la clasificación mundial, la educación secundaria se equipara a la enseñanza media, esto es, incluye lo que en México se denomina secundaria y bachillerato.

⁴ Acaso ninguna otra ciencia de la educación, como recordó King, pueda permitirse el lujo de ser tan fiel a sus orígenes históricos y a la tradición como la educación comparada. Nacida en los albores del siglo XIX, en el preciso momento en que se empezaron a plasmar los primeros sistemas escolares públicos, su objetivo ha seguido siendo, invariablemente, el mismo: el estudio de los sistemas educativos extranjeros y su comparación con vistas a contribuir a la solución de problemas en la esfera pedagógica mediante el examen de las soluciones aportadas en otros contextos, particularmente en otros países (Pedró, 1993:21).

⁵ Por esta razón, en el examen metropolitana que se aplica en la zona metropolitana de la ciudad de México se afirma que aquellos estudiantes que tengan por lo menos 32 aciertos (de 128) tendrán garantizado un lugar en alguna institución pública de esta zona.

⁶ En la actualidad todas las instituciones de enseñanza media y superior aplican un examen de ingreso. Esta práctica se fue generalizando conforme la relación oferta/demanda se fue distanciando. Varias instituciones aplican su examen propio, tal es el caso de las universidades Nacional Autónoma de México, Veracruzana y Autónoma de Baja California. Las dos últimas ofrecen este servicio a otras instituciones en el nuevo mercado de exámenes de ingreso que se ha establecido desde la década de los noventa; otras, como instrumento de ingreso, aplican el examen del CENEVAL.

⁷ “Cuando se diseñó el IDANIS la organización de los contenidos en los programas de estudio de educación primaria y secundaria respondía a un modelo basado en la estructura de las disciplinas. Este modelo, llamado por algunos autores *conceptual-empírico* [...] el IDANIS fue concebido a partir de un marco distinto: no se consideró necesario medir el dominio que tuvieran los alumnos sobre los contenidos disciplinares partiendo de unidades atomizadas. Por el contrario se abordó la medición de la aptitud para el aprendizaje [...]. En la paulatina aceptación del IDANIS tuvieron un papel importante los siguientes factores: primero, las adecuaciones que ciertamente se incorporaron a los programas de estudio anteriores a 1993, durante el periodo de su vigencia, las cuales disminuyeron la rigidez de los mismos y significaron en ocasiones replanteamientos de enfoque acerca de temas concretos; segundo, la actualización de una importante cantidad de profesores de primaria y secundaria a través de programas institucionales y de manera independiente, lo que les permitió un contacto con nuevas posiciones ante la enseñanza; tercero, la mayor difusión de nuevos enfoques educativos a través de libros, revistas y eventos académicos; y, cuarto, la presencia de corrientes de opinión entre la comunidad educativa y la sociedad en general, en favor de una enseñanza que permitiera a los alumnos la solución de situaciones-problema, en oposición a un tipo de enseñanza enciclopédica de corte memorístico” Tomado de: Dirección General de Evaluación Educativa (disponible en: http://snec.sep.gob.mx/Documentos/Ensayos/EnsayoIdanis1/IDANIS_Contexto.htm).

⁸ Según el artículo 2 de su Estatuto: “El CENEVAL tiene por objeto contribuir a mejorar la calidad de la educación media superior y superior mediante evaluaciones externas de los aprendizajes logrados” y el inciso 2 del artículo tercero establece como directriz general “Proceder conforme a mecanismos y métodos que aseguren la confiabilidad y confidencialidad de la información, el profesionalismo de los evaluadores, la pertinencia de los procedimientos, la seguridad de los bancos de reactivos y los exámenes, la imparcialidad de la calificación [...] y el inciso 3: “difundir públicamente los resultados de las evaluaciones” *Estatuto*, Centro Nacional de Evaluación. Con las reformas aprobadas hasta 2004 (disponible en http://www.e-mexico.gob.mx/wb2/eMex/eMex_CENEVAL_apren).

⁹ Así lo consigna la propia Dirección General de Evaluación “Durante este periodo ganó terreno la idea de que los resultados no debían de ser difundidos [...] de suerte los resultados de evaluación se convirtieron en información confidencial. Esta política de silencio obligado trajo consigo muchos inconvenientes [...]” (SEP, 2005:3).

¹⁰ Esta visión aún la conserva la información de la Dirección General de Evaluación, pues no es fácil entrar a la base de datos que tiene en la página electrónica de la SEP. El periódico *Reforma* publicó una serie de reportajes en la semana del 9 de enero de 2005 ya que, de alguna forma, obtuvo el *password* que le permitió acceder a información que se encuentra restringida. Los títulos de los reportajes hablan por sí mismos: “Caen escuelas del DF en aprovechamiento” (Sonia del Valle), “Lo que necesito es un mejor salario” (Pedro Flores que entrevista al profesor mejor evaluado a nivel nacional en 2004), “Cambian a director a pesar de sus logros”, “Concentran Puebla, Chiapas y Guanajuato el mayor número de escuelas primarias con nivel más bajo del país”, “Truenan matemáticas los mejores alumnos”. Aunque la SEP protestó en el mismo periódico por la interpretación que se hacía de la información obtenida, lo que es cierto, es que ésta sólo se encuentra al alcance de los funcionarios de la Dirección General y de la SEP. Ello confirma que la evaluación en México no se realiza para retroalimentar al sistema ni tiene un

carácter formativo. Los medios de comunicación dan un sentido a esta información, pero los maestros y alumnos, que son los que pueden modificar su trabajo educativo, no son sujetos de estos datos (*Reforma*, reportajes sobre resultados de la evaluación, publicados en la semana del 9 al 15 de enero de 2005).

¹¹ Así, el CENEVAL está publicando reportes académicos en formatos de libro como el de Tirado (2004) *Evaluación de la educación en México. Indicadores del EXANI-III*, así como *La inteligencia colectiva de México* (2005), o las publicaciones que mensualmente edita en la revista *Este País*, con nombres por demás sugerentes: “Factores que influyen en los resultados educativos: el caso del Metropolitano en 2003”, “Examen general para el egreso de la licenciatura en ingeniería química”; “Exámenes generales: ingeniería eléctrica e ingeniería electrónica”, así como un boletín bimestral.

¹² Aunque como otro rasgo de la compulsión a la evaluación que caracteriza a la política educativa nacional, subsiste junto con el INEE, la Dirección General de Evaluación, responsable de elaborar, aplicar y procesar varias pruebas nacionales: IDANIS, Factor de Aprovechamiento Escolar, Preparación Profesional. México es un país donde existen doubles instancias de evaluación. Cfr. Díaz Barriga, A. (2005).

¹³ Entre algunos de los documentos que se encuentran a disposición del público en la página web del Instituto se encuentran los siguientes: *Panorama educativo de México. Indicadores del sistema educativo nacional 2004*; *La calidad de la educación básica en México 2004*; *Resultados de la prueba PISA 2000 y 2003 en México*; y *Cómo comprender los resultados de México en la prueba PISA 2003*.

¹⁴ Cfr. *Informe de actividades del INEE*.

¹⁵ El reporte del Laboratorio de la Calidad de la Educación para América Latina y el Caribe del año 1998, menciona que el examen se aplicó a diversos países de la región, pero que uno pidió que sus resultados se retiraran del informe que se publicaría. Lo que no indica el documento es que ese país fue México, obviamente los resultados de sus estudiantes no eran satisfactorios. Cfr. LLECE-OREALC-UNESCO (1998); ver notas de los periódicos “Esconden califica-

ción reprobatoria” (*Reforma*, núm. 2862, año 8, sección A, lunes 15 de octubre de 2001), “Consideré mi obligación utilizar la información disponible de manera responsable... hubiera sido negativo y pernicioso para la cultura de la evaluación que los resultados hubieran sido publicados”, en Reyes, C. Melgar, I, Guerrero, C. “Oculto Limón datos para no desanimar”, *Reforma*, núm. 2862, año 8, sección A, lunes 15 de octubre de 2001.

¹⁶ Por demás está afirmar que esta falta de cuidado ante el manejo técnico de los resultados de la información se observaron con estupefacto de la población en el año 2004, cuando el Presidente de la república declaraba que tenía una gran noticia que dar “los estudiantes mexicanos son los mejores de América Latina”, para que al día siguiente un funcionario de la OCDE declarara que México no tenía nada de qué enorgullecerse, que los resultados de los estudiantes mexicanos estaban muy lejos de los desempeños que habían obtenido los de otros países.

¹⁷ Aun cuando se revisen diversos documentos, por ejemplo “Resultados de la secundaria: globales, de razonamiento verbal y matemático”, “Resultados de la secundaria por área de conocimiento y régimen de escuela de procedencia”

o los mismos documentos de formulación del examen, no se identifica una mayor fundamentación de esta cuestión. El que contiene una mayor referencia al examen, aunque muy indirecta a los supuestos que orientan la selección de su contenido es: CENEVAL (2004a) *EXANI-I. Guía de examen, 2004*.

¹⁸ No puedo entrar en este espacio a analizar las deficiencias metodológicas que se pueden encontrar en los instrumentos de contexto, que son evidentes para cualquiera que conozca las condiciones de trabajo de las secundarias en México. Sencillamente, deseo resaltar la importancia que tiene que estos aspectos sean tomados en cuenta en la elaboración del informe final.

¹⁹ Únicamente tomemos en cuenta que con la bandera de “abajo el teoricismo”, en la última reforma de la educación normal no sólo se eliminaron las materias eminentemente teóricas: teoría, filosofía e historia de la educación. Sino que se eliminaron un conjunto de materias instrumentales que tienen la finalidad de apoyar la formación profesional del estudiante, tales como didáctica, psicología infantil y/o adolescencia, del aprendizaje, grupal, entre otras. Hoy se forman profesionales de la educación con una perspectiva empirista de su trabajo.

Bibliografía

- Anagnostopoulos, D. (2005). “Testint, test and classroom texts”, *Journal of Curriculum Studies*, vol 37, núm. 1 (Taylor & Francis Ltd), p. 36. (disponible en: <http://www.tandf.co.uk/journals>).
- Banco Mundial (1992). *Educación primaria. Documento de política*, Washington: Banco Mundial.
- Teobaldo, M.; Bertoni, A. y Poggi, M. (1995). *Evaluación: nuevos significados para una práctica compleja*, Buenos Aires, Kapelusz.
- CENEVAL (2004a). *EXANI-I Guía de examen 2004. Examen Nacional de Ingreso a la Educación Media Superior*, México: CENEVAL.
- CENEVAL (2004b). *Cómo prepararse para el Examen Nacional de Ingreso a la Educación Media Superior 2004. EXANI-I*, México: CENEVAL.
- Cramer, J. (1983). “Merit Pay: Challenge of th decade”, *Curriculum Review*, diciembre, pp. 7-10.
- Díaz Barriga, A. (2005). “Los riesgos de los programas de evaluación y acreditación de la educación superior”, ponencia presentada en Seminario regional Las nuevas tendencias de la evaluación y de la acreditación en América Latina y el Caribe, Buenos Aires: IESALCF-UNESCO-CONEAU, 7 de junio.

- EURYDICE (1997). *Una década de reformas en la educación obligatoria de la Unión Europea (1984- 1994)*. La Red Europea de información en Educación.
- EURYDICE (2002). *Competencias clave. Un concepto en expansión dentro de la educación general obligatoria*, La Red Europea de información en Educación.
- LLECE-OREALC-UNESCO (1998). *Primer estudio internacional comparativo sobre lenguaje, matemáticas y factores asociados en tercero y cuarto grado*. Santiago de Chile: Laboratorio Latinoamericano de Evaluación de la Calidad Educativa-OREALC-UNESCO (disponible en <http://www.unesco.cl/programa/inflab1.pdf>).
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*, Madrid, Pirámide.
- OCDE (2001). *Conocimientos y aptitudes para la vida. Resultado de PISA 2000*, col. Aula XXI, México: OCDE/Santillana.
- OCDE (2002). *Muestra de reactivos empleados en la evaluación PISA 2000. Aptitudes para la lectura, matemáticas y ciencias*, col. Aula XXI, México: OCDE/Santillana..
- OCDE (2004). *Marcos teóricos de PISA 2003 Conocimientos y destrezas matemáticas, lectura, ciencias y solución de problemas*, OCDE-INECSE/Ministerio de Educación y Cultura de España, documento recuperado de la página web del Ministerio de Educación y Cultura de España.
- Pedró, F. (1993). “Conceptos alternativos y debates teórico-metodológicos en educación comparada: una panorámica introductoria”, en Schriewer y Pedró (eds.), *Manual de educación comparada*, Barcelona: Promoción y Publicaciones Universitarias.
- SEP (2002). *La experiencia de evaluación en la Dirección General de Evaluación. Treinta años de experiencia en medición del logro educativo en educación básica y normal* (disponible en: <http://snec.sep.gob.mx/>).
- Stuttlebeam, D. y Shinkfield, A. (1987). *Evaluación sistémica*, serie Temas de educación, Barcelona: Paidós.
- Tirado, F. (2004) *Evaluación de la educación en México. Indicadores del EXANI-I*, México: Centro Nacional de Evaluación.

Artículo recibido: 17 de junio de 2005

Dictamen: 4 de octubre de 2005

Aceptado: 12 de octubre de 2005