

Exámenes de la Calidad y el Logro Educativos (Excale): nueva generación de pruebas nacionales

Eduardo Backhoff, Lucía Monroy,
Margarita Peón, Andrés Sánchez y
Ma. de Lourdes Tanamachi*

CUADERNO No. 17



Instituto Nacional para la
Evaluación de la Educación

**COLECCIÓN CUADERNOS
DE INVESTIGACIÓN**

ISSN 1665-9457

*Instituto Nacional para la Evaluación de la Educación

Este texto puede consultarse en: www.inee.edu.mx

MÉXICO, NOVIEMBRE, 2005

CONTENIDO

Resumen	3
Introducción	4
El contexto y el aprendizaje escolar	6
Propósitos y principios de la evaluación del aprendizaje	9
Proceso de diseño, construcción, aplicación y validación de los Excale	11
Plan de evaluación del aprendizaje	19
Conclusiones	20
Referencias bibliográficas	23
Anexo: Programas evaluativos de la educación básica a cargo de la DGE	26

RESUMEN

En el presente trabajo se reseña la puesta en marcha de una nueva generación de pruebas nacionales que el Instituto Nacional para la Evaluación de la Educación (INEE) utiliza a partir de junio de 2005, en muestras de estudiantes con representatividad nacional y estatal, para evaluar la calidad del Sistema Educativo Nacional, con base en los aprendizajes que logran los estudiantes. Se describe el modelo de logro educativo adoptado por el INEE para el diseño de sus evaluaciones, así como los propósitos y principios de la evaluación del aprendizaje que asume este Instituto. Con mayor detalle, se explica el proceso para diseñar, construir y validar los Exámenes de la Calidad y el Logro Educativo (Excale), que se definen como pruebas de tipo criterial, alineadas al currículum y con un diseño matricial. A manera de conclusión, se discute la necesidad de contar con instrumentos evaluativos de calidad para poder cumplir con las metas del INEE.

Palabras clave: evaluación educativa, calidad educativa, pruebas nacionales, educación básica, exámenes de logro escolar.

Hasta el año de 2002 la Dirección General de Evaluación (DGE), de la Secretaría de Educación Pública (SEP), fue la única instancia gubernamental responsable de evaluar al Sistema Educativo Nacional (SEN). Para ello, desde su creación en la década de los años setenta, la DGE diseñó y operó un sinnúmero de programas e instrumentos evaluativos con el fin de conocer y mejorar la calidad de la educación básica (en el anexo se muestran los programas evaluativos más importantes de la DGE, de los últimos años).

Aunque no todos los programas han tenido el impacto deseado, la DGE logró poner en la mesa de las decisiones políticas el tema de la evaluación como vía indispensable para mejorar la calidad de la educación nacional. En consecuencia, el Gobierno Federal creó el Programa Nacional de Educación 2001-2006 (Pronae), el cual contempló la creación de mecanismos rigurosos y confiables de evaluación que operen con independencia de las autoridades para permitir rendir cuentas a la sociedad (SEP, 2001).

El Pronae estableció dos grandes metas: 1) la creación del Instituto Nacional para la Evaluación de la Educación (INEE), y 2) la redefinición de las funciones de la DGE. El primero se concibió como un organismo con independencia y autonomía, cuyo objetivo primordial es conocer y explicar la calidad del SEN en su conjunto, así como el de dar a conocer los resultados de sus evaluaciones a las autoridades educativas y a la sociedad en general. Por su parte, la DGE se redefinió como un organismo que forma parte de la Administración Central, cuyo propósito es realizar evaluaciones diagnósticas y formativas, orientadas a retroalimentar la toma de decisiones sobre individuos e instituciones (DGE, 2004).

Con la idea de evaluar la calidad de la educación en México, la DGE desarrolló desde hace más de un lustro lo que ahora se conoce como las Pruebas de Estándares Nacionales (PEN). Dichos instrumentos fueron utilizados durante el periodo 1998-2003, con muestras representativas de escolares de todo el país. Sin embargo, al crearse el INEE, las PEN pasaron a ser parte de su responsabilidad; función que asumió de manera parcial con el análisis de los resultados de la aplicación realizada –todavía por la DGE– en junio de 2003. Dichos resultados fueron publicados en su primer informe *La Calidad de la Educación Básica en México* (INEE, 2003).

Siguiendo el mismo esquema metodológico propuesto por la DGE en cuanto a estructura y contenidos, el INEE elaboró una nueva versión de las PEN, las cuales aplicó en junio de 2004 con una muestra representativa de alumnos de todo el país, cuyos resultados se publicaron en su segundo informe anual *La Calidad de la Educación Básica en México: resultados de evaluación educativa 2004* (INEE, 2004a).

Las PEN se diseñaron con la idea de evaluar las habilidades cognitivas más que los conocimientos curriculares. Para ello se desarrollaron varias pruebas de Comprensión Lectora y Matemáticas, dirigidas a estudiantes de primaria y secundaria. En el primer caso las pruebas se construyeron por grado escolar (2º, 3º, 4º, 5º y 6º). Para secundaria se construyó una sola prueba dirigida a estudiantes de los tres grados. Los reactivos de las PEN se calibran

con base en el modelo de Rasch (Wright & Stone, 1998), para después transformar las puntuaciones a una escala con un rango de 200-800 unidades, con una media de 500 y una desviación estándar de 100. Hasta 2002, la calibración se hizo considerando ambas áreas temáticas (Comprensión Lectora y Matemáticas) en una sola escala (Backhoff, Andrade, Peón, Sánchez, Juárez, Monroy y Tanamachi, 2003). A partir de 2003, ya bajo la responsabilidad del INEE, la calibración se realizó por separado, conformando dos escalas: Matemáticas y Comprensión Lectora.

Desafortunadamente, el diseño de las PEN tiene deficiencias técnicas importantes, por lo cual no es recomendable realizar comparaciones de los resultados en el tiempo y, por consiguiente, no es posible conocer las tendencias en el aprendizaje de los estudiantes. Debido a lo anterior, es difícil comparar en forma confiable los resultados de dos aplicaciones, como lo han venido demandando varios sectores de la sociedad (Martínez-Rizo, 2004; Backhoff y Martínez-Rizo, 2004). A estas deficiencias se suma el hecho de que la documentación en donde se plasma el marco teórico, o de referencia, que sustenta las pruebas es escasa, por lo cual no queda claro si se realizaron los procesos mínimos que validan estos instrumentos. Adicionalmente, las PEN no se diseñaron estrictamente con el propósito de evaluar los contenidos de los planes y programas de estudios nacionales; por ese motivo revelan poco acerca del dominio de los contenidos curriculares que poseen los escolares, y con esto se limita la información que se le puede dar a las autoridades educativas respecto al logro de los propósitos del currículum mexicano (Backhoff y Martínez-Rizo, 2004; Backhoff, 2005).

Estas limitaciones son importantes si se toman en cuenta los siguientes propósitos del INEE relacionados con la evaluación del aprendizaje: 1) construir una visión general de lo que los estudiantes aprenden como resultado de su escolarización formal; 2) conocer los puntos fuertes y débiles del aprendizaje de los estudiantes en las asignaturas de mayor importancia, y 3) permitir las comparaciones del rendimiento escolar, así como las tendencias de aprendizaje a lo largo del tiempo.

Teniendo en cuenta las limitaciones de las PEN para lograr estos propósitos, el Consejo Técnico del INEE optó por elaborar un Plan General de Evaluación del Aprendizaje, en el cual se contempla la construcción de una nueva generación de pruebas nacionales: los Exámenes de la Calidad y del Logro Educativos (Excale). La construcción de estos exámenes se inició en febrero de 2004 y su primera aplicación a nivel nacional se realizó en junio de 2005 (INEE, 2004b).

A partir de lo anterior, los objetivos de este trabajo son: 1) describir el modelo de logro educativo utilizado en estas pruebas; 2) exponer los propósitos y principios del INEE en relación a la evaluación del aprendizaje; 3) describir en forma resumida el proceso de diseño, construcción y validación de las pruebas Excale; 4) presentar el plan de pruebas para los próximos años, y 5) discutir la importancia de contar con instrumentos robustos para poder evaluar la calidad de los servicios educativos que se ofrecen en el país.

El contexto y el aprendizaje escolar

Además de describir los niveles de logro que alcanzan los alumnos del SEN, con la aplicación de sus pruebas, el INEE trata de aportar elementos para explicar las diferencias en el rendimiento académico de los alumnos, considerando las variables de contexto en que ocurre el aprendizaje (Backhoff, Monroy y Tanamachi, 2005).

En la actualidad se cuenta con un vasto acervo de investigaciones empíricas que muestra la complejidad de variables y relaciones potenciales que influyen en los resultados de aprendizaje de los estudiantes (Scheerens, 1997). Este campo se ha estudiado desde distintas perspectivas teóricas que se entrecruzan –teorías psicológicas y cognitivas, sociológicas y antropológicas, pedagógicas y curriculares– las cuales aportan componentes que ayudan a tener una idea general del tema (OCDE, 2004).

Por lo anterior, el INEE no se propuso elaborar un nuevo modelo teórico que agregaría poco a los muchos modelos que se encuentran disponibles en la literatura. En su lugar, la intención fue sacar ventaja de los modelos existentes sobre la teoría curricular y el aprendizaje escolar. Al respecto, partimos de que la mayoría de los autores coinciden en que el currículum tiene tres niveles, los cuales se ven influidos por factores contextuales, institucionales y personales (Mullins *et al.* 2003). En el primer nivel se ubica el *currículum formal o intencional* que se diseña desde las cúpulas administrativas del sistema educativo. El *currículum implementado* es el segundo nivel, y resulta de la interpretación que cada maestro hace del currículo formal, considerando lo que es apropiado para un grupo particular de estudiantes. Finalmente, el tercer nivel es el *currículum logrado*, el cual se refiere al grado en que los estudiantes aprenden de las actividades que organizan los maestros en clase para tal propósito.

La figura 1 muestra el modelo adoptado por PISA (por sus siglas en inglés *Programme for International Students Assessment*), representando las variables potenciales que inciden en el aprendizaje escolar –propuesto inicialmente por Carroll (1963), modificado por la *International Evaluation Association* (IEA) y enriquecido por la OCDE–, esquema que el INEE adoptó para el caso particular de sus evaluaciones. Las flechas indican la dirección de las relaciones y, aunque pueda haber relaciones recíprocas y entrecruzadas, solamente se señalan las de mayor intensidad y obviedad (OCDE, 2004).

El modelo de PISA está orientado a explicar las *oportunidades para aprender*, como la base fundamental del aprendizaje del estudiante. Asimismo, comparte con otros modelos aspectos básicos, como la distinción entre niveles de sistema, escuela, salón de clases y estudiantes, así como entre antecedentes o insumos (lo que se espera que aprenda el estudiante), procesos (cómo se organiza la instrucción) y resultados (lo que el estudiante aprende). El modelo intenta distinguir entre el currículum formal o intencionado (a nivel del sistema), el implementado (a nivel del salón de clases) y el logrado (a nivel del estudiante).

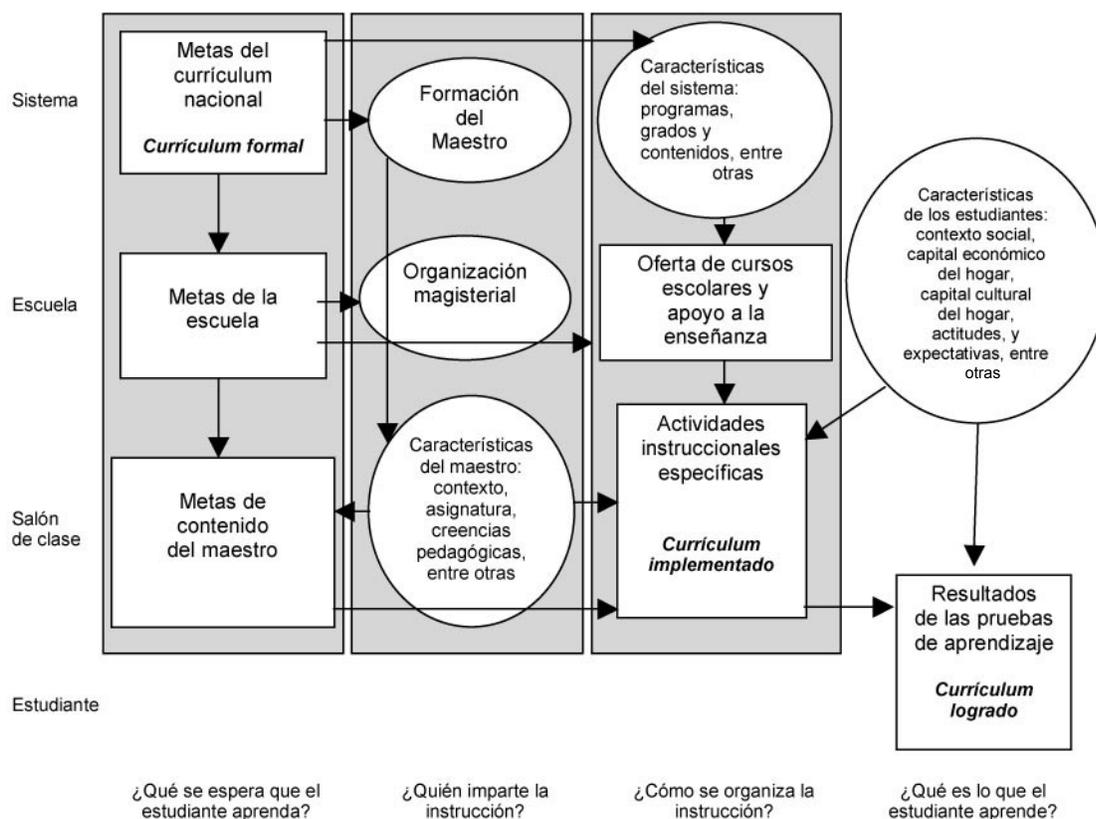


Figura 1. Modelo de aprendizaje escolar de PISA (Fuente: OCDE, 2004)

A partir de este modelo de aprendizaje escolar, el INEE elaboró su propio modelo de logro educativo, adaptándolo a las características y propósitos de las pruebas nacionales. Dado que se asume que las oportunidades de aprendizaje son un elemento crucial para los resultados académicos de los alumnos, el currículum cobra un papel central en el modelo, ya que a través de él se decide qué deberán aprender, cómo deben hacerlo y en qué orden se enseñarán los contenidos (véase figura 2).

Por tal motivo, el modelo de logro educativo propone como *eje organizador al currículum*, entendiéndolo como la norma oficial mexicana que estructura y delimita lo que se desea que los alumnos aprendan en cada una de las asignaturas de los diferentes grados escolares. Intenciones que se traducen en secuencias de oportunidades para el aprendizaje proporcionadas por el maestro, con el fin de que los estudiantes adquieran los conocimientos y habilidades especificados en los contenidos del currículum. Por ello, la definición del currículum del Modelo de Logro Educativo del INEE, incorpora niveles o aspectos de éste que van más allá del solo planteamiento de intenciones, pues integra las transformaciones

que puede tener el currículum oficial a lo largo del proceso educativo hasta ser traducido en el aprendizaje logrado por los alumnos.

En este modelo se asume que el proceso educativo está anidado por diferentes niveles: el alumno, el salón de clase, la escuela, la entidad federativa y el Sistema Educativo Nacional. Cada uno de estos niveles tiene características específicas que influyen en los productos requeridos para que fluya el proceso de educación, los cuales se traducen en los cinco aspectos del currículum, a saber: el currículum oficial, el currículum adaptado por la entidad, el currículum enriquecido por la escuela, el currículum implementado por el maestro y, finalmente, el currículum logrado que se conforma con los conocimientos y habilidades adquiridos por el alumno como producto de las oportunidades de aprendizaje a las que fue expuesto.

En la figura 2 se muestran de manera gráfica los niveles de agregación, así como las variables que influyen en el logro educativo. Con la finalidad de contar con un modelo parsimonioso no se presenta la totalidad de las relaciones interactivas que pudiesen existir entre las variables.

Este Modelo de Logro Educativo sirve de base para elaborar los cuestionarios de contexto que buscan identificar las variables explicativas de las diferencias en el logro del aprendizaje que muestran los distintos sectores estudiantiles a evaluar, por ejemplo los estratos o modalidades educativos: escuelas públicas urbanas, indígenas, privadas, entre otros.

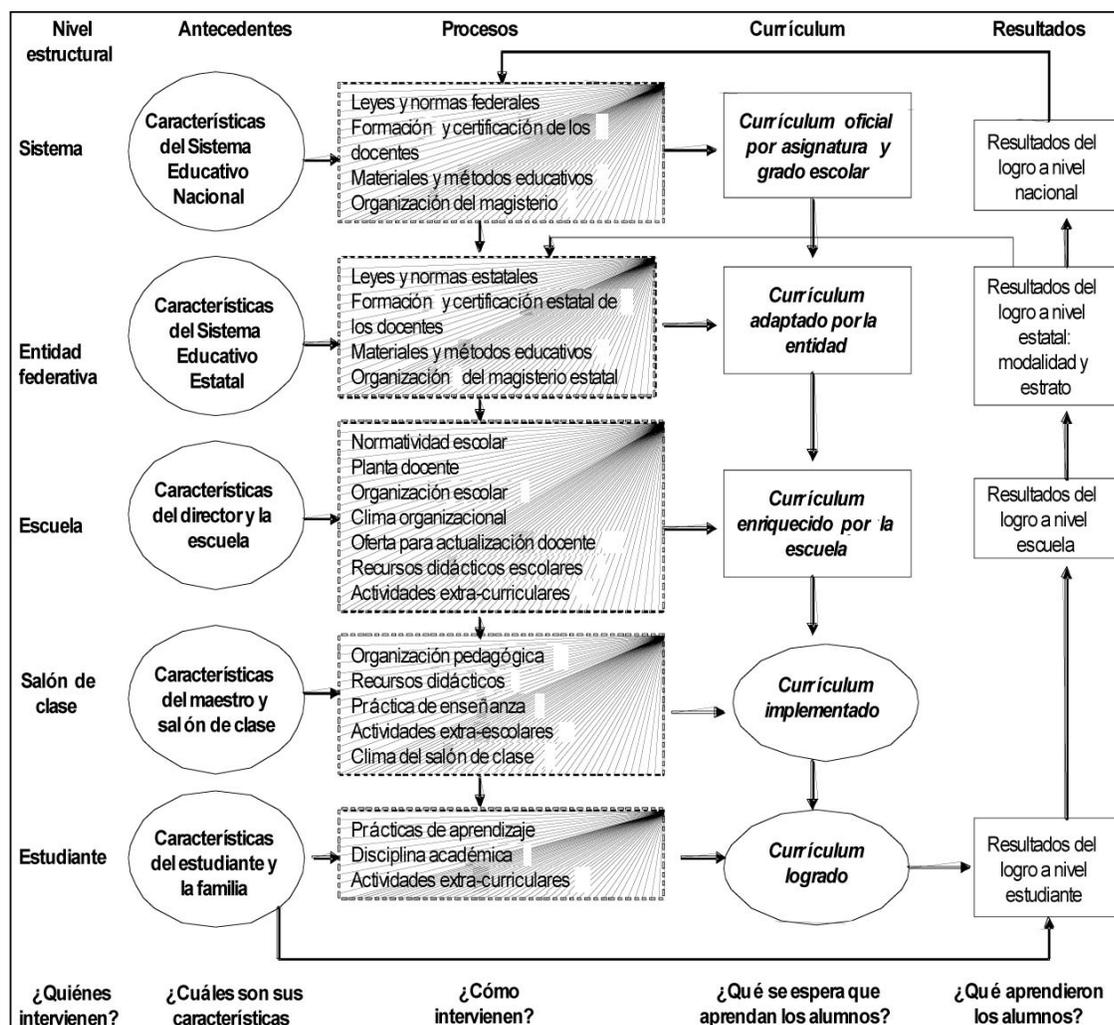


Figura 2. Modelo de Logro Educativo del INEE, donde las variables latentes (no observadas directamente) se muestran en óvalos; en rectángulos sólidos se muestran las variables manifiestas. Para facilitar el esquema, se presentan en rectángulos punteados las variables que combinan indicadores observables con indicadores latentes.

Propósitos y principios de la evaluación del aprendizaje

La evaluación del aprendizaje que realiza el INEE tiene como fin proporcionar un conocimiento general del rendimiento académico de los estudiantes de educación básica y media superior, a niveles estatal y nacional, con distintos grados de desagregación por estrato y modalidad educativos (INEE, 2004b). Entre los propósitos de mayor relevancia que persigue tal evaluación, se encuentran los siguientes:

- ◆ Conocer el logro académico de los estudiantes a niveles estatal y nacional, así como los factores de contexto más importantes que explican las diferencias de los sectores estudiados.
- ◆ Contribuir al conocimiento de los alcances y limitaciones del Sistema Educativo Nacional y, con ello, el nivel de la calidad de la educación básica en México.
- ◆ Emitir juicios de valor contextualizados que sirvan para apoyar una toma de decisiones documentada.
- ◆ Complementar los procesos evaluativos existentes, desarrollados por otras instancias nacionales e internacionales (por ejemplo la DGE y la OCDE).
- ◆ Aportar elementos para enriquecer la rendición de cuentas a la cual tiene derecho la sociedad mexicana.

Asimismo, el INEE ha establecido que la realización de evaluaciones del aprendizaje a gran escala debe partir de los siguientes principios:

- ◆ Ser una evaluación externa del SEN del país.
- ◆ Ser de alta calidad y apegarse a normas y prácticas reconocidas internacionalmente.
- ◆ Proporcionar información que dé una imagen válida y confiable del SEN en su conjunto.
- ◆ Asegurar que las evaluaciones respeten el valor de equidad, en particular las relacionadas con las diferencias de género y de grupos étnicos.
- ◆ Ser realizadas de manera clara y transparente, con la participación colegiada de maestros y especialistas.
- ◆ Estar legitimadas académica y socialmente.
- ◆ Proporcionar elementos que ayuden a mejorar la calidad del sistema educativo del país.

Para lograr los propósitos de la evaluación del aprendizaje, considerando las premisas del INEE, el Consejo Técnico de este Instituto definió que sus pruebas Excale tuvieran tres características básicas: 1) ser de tipo criterial; 2) estar alineadas al currículum nacional, y 3) tener un diseño matricial.

Se entienden como *pruebas criterios* aquellas que se diseñan para conocer con precisión el grado de dominio del estudiante sobre un conjunto de contenidos específicos. El referente para interpretar los resultados de este tipo de pruebas es la cantidad y tipo de material que el estudiante maneja del universo de contenidos evaluados, o constructo de la prueba (Popham, 1990). Por lo general, las pruebas criterios se utilizan tanto para diagnosticar como para certificar la posesión de conocimientos; a diferencia de las pruebas normativas, que se utilizan con propósitos de selección de individuos, y donde se pone menor énfasis en la descripción detallada de los contenidos que se evalúan.

Por otro lado, las *pruebas alineadas* al currículum son aquellas que se elaboran con una metodología *ad hoc* para evaluar con precisión los contenidos curriculares (llámense habilidades o conocimientos), los cuales se definen en los planes y programas de estudio

(por ejemplo: comprensión de lectura en 3º de primaria). El alineamiento curricular de las pruebas implica: identificar los resultados importantes pretendidos por el currículo, asociar las acciones de evaluación con los contenidos esenciales, definir el dominio curricular completo sobre el cual se desarrollará el examen y precisar los resultados de aprendizaje que se establecen en el currículum oficial (Nitko, 1994). México cuenta con un currículum nacional, libros de textos gratuitos y una formación magisterial uniforme, condiciones que lo convierten en un candidato idóneo para este tipo de pruebas; a diferencia de otros países (como Estados Unidos), quienes para evaluar la educación a nivel nacional tienen que darse a la enorme tarea de generar *estándares de contenido y ejecución* en cada disciplina, con el fin de alinear sus pruebas a éstos.

Finalmente, las *pruebas de tipo matricial* son aquellas diseñadas para evaluar una gran cantidad de contenidos, sin someter a los estudiantes a largas jornadas de resolución de pruebas. Para ello se requiere construir un conjunto de reactivos que cubran el dominio completo que se desea evaluar, y después dividirlo en subconjuntos para repartirlos entre los estudiantes, de tal manera que cada alumno conteste sólo algunas preguntas (Deng, Ferriss, & Hombo, 2003; Van der Linden, Veldkamp, & Carlson, 2004). Con un diseño matricial se limita el número de reactivos que contesta cada estudiante (reduciendo con ello el tiempo de la evaluación), a la vez que se cubre la totalidad de contenidos curriculares seleccionados entre todos los estudiantes. Como consecuencia, la calificación individual del estudiante pierde precisión –en tanto que su error de medida aumenta–, no así los resultados agregados a nivel de entidad federativa y modalidad educativa, que son el centro de interés del INEE (Gaviria, 2005).

Proceso de diseño, construcción, aplicación y validación de los Excale

Es importante mencionar que para poder diseñar pruebas nacionales de calidad se requiere seguir los lineamientos y estándares precisos para la construcción de instrumentos de evaluación educativa (AERA, APA & NCME, 1999; Martínez-Rizo y col., 2000). Asimismo, es importante adoptar una metodología rigurosa que haya probado su validez para los propósitos que se persiguen; éste es el caso del modelo de construcción de pruebas criterios alineadas al currículum, donde el aspecto central es el trabajo colegiado de especialistas y docentes. Por lo anterior, el INEE adoptó para las pruebas Excale dicho modelo (véase Nitko, 1994), adaptándolo a las necesidades nacionales.

Así, el proceso quedó definido en siete fases y 16 etapas básicas. La tabla I muestra el proceso de diseño, construcción y validación de las Pruebas Nacionales del INEE. En ella se aprecian con detalle los procedimientos y productos que se esperan de cada etapa, así como el personal externo al INEE participante en cada una de ellas. Es importante señalar que, por lo general, los productos de cada etapa sirven como insumos de la siguiente, por

lo que en el propio proceso de generación de este tipo de pruebas se contempla, en parte, el proceso de validación de las mismas (Contreras, 2000; Contreras, Backhoff y Larrazolo, 2003).

Como se observa en esa tabla, a lo largo de todo el proceso participan diversos especialistas, grupos de asesores, comités de expertos y personal técnico de la Dirección de Pruebas y Medición. Asimismo, en cada etapa del proceso se utilizan diversos procedimientos, entre los que destacan: 1) la documentación de procesos similares para la construcción de pruebas de aprendizaje a gran escala, realizados por instituciones de reconocida calidad internacional; 2) la capacitación dirigida a los cinco comités de especialistas y docentes que participan en el proceso; 3) la elaboración y preparación de materiales para el trabajo de los cinco comités; 4) el trabajo colegiado, donde se toman decisiones de mayor importancia (por ejemplo, el tipo de diseño matricial); y 5) el trabajo individual por encargo, que requiere la experiencia y conocimientos de especialistas en un tema en particular (por ejemplo, el diseño de la muestra poblacional).

Tabla I. Proceso de diseño, construcción, aplicación y validación de los Excale

Fases	Etapas	Participantes externos ¹	Procedimientos	Productos
I. Planeación general ²	1. Diseño del Plan General de Evaluación	<ul style="list-style-type: none"> ◆ Consejo Técnico ◆ Asesores en medición y validación 	1. Documentación 2. Seminarios 3. Trabajo colegiado 4. Trabajos por encargo	1. Plan General de Evaluación del Aprendizaje, con el marco de referencia de los Excale 2. Manual general de procedimientos 3. Manual técnico para el diseño matricial 4. Manual técnico para el escalamiento de puntuaciones y niveles de logro 5. Marco teórico de validación de los Excale
	2. Diseño y elaboración de cuestionarios de contexto	<ul style="list-style-type: none"> ◆ Especialistas en diseño de cuestionarios ◆ Especialistas en evaluación del aprendizaje 		6. Marco de referencia de los cuestionarios de contexto 7. Cuestionarios de contexto de alumno, docente y director
	3. Diseño y desarrollo del sistema informático	<ul style="list-style-type: none"> ◆ Especialistas en bases de datos 		8. Sistema de bases de datos de reactivos 9. Documento que describe la estructura y funcionamiento de la base de datos ³

¹ Personal de la Dirección de Pruebas y Medición interviene en todas las etapas del proceso.

² Fase general para la nueva generación de las pruebas Excale.

³ A partir de la etapa cuatro, el sistema informático se irá alimentando con la información que se produzca a lo largo de todo el proceso.

Fases	Etapas	Participantes externos	Procedimientos	Productos
II. Estructuración de los Excale	4. Diseño de las pruebas	◆ Comités académicos (uno por prueba)	1. Documentación 2. Capacitación 3. Preparación de materiales 4. Trabajo colegiado	10. Manual técnico para el diseño de pruebas nacionales 11. Reticula curricular de cada prueba 12. Tabla de contenidos de cada prueba 13. Manual técnico para la elaboración de especificaciones 14. Especificaciones de reactivos de cada prueba 15. Dos revisiones por especificación de reactivos
	5. Especificación de reactivos	◆ Comités elaboradores de especificaciones de reactivos (uno por prueba)		
III. Construcción de reactivos de los Excale	6. Elaboración de reactivos	◆ Comités constructores de reactivos (uno por prueba)	1. Documentación 2. Capacitación 3. Trabajo individual y colegiado	16. Manual técnico para la construcción de reactivos 17. Tres reactivos por especificación 18. Dos revisiones por reactivo construido
	7. Validación de reactivo	◆ Comités de validación y sesgo (uno por prueba)		
	8. Piloteo de reactivos y cuestionarios de contexto	◆ Algunas coordinaciones estatales de evaluación		1. Muestreo para el piloteo 2. Preedición de reactivos e impresión de cuadernillos 3. Capacitación
IV. Conformación de los Excale	9. Selección de reactivos e integración de bloques y formas	◆ Asesores en medición	1. Documentación 2. Análisis de reactivos	26. Manual técnico para el análisis psicométrico de reactivos 27. Informe de estimadores estadísticos de reactivos 28. Bloques de reactivos 29. Estructura de formas (combinación de bloques)
	10. Edición, armado e impresión	◆ Imprenta contratada		

Fases	Etapas	Participantes externos	Procedimientos	Productos
V. Aplicación de los Excale	11. Muestreo poblacional	◆ Especialistas en muestreo	1. Documentación 2. Trabajo colegiado 3. Trabajo por contrato	33. Manual técnico para seleccionar muestras poblacionales 34. Marco muestral actualizado 35. Diseño y muestra poblacional
	12. Administración de las pruebas y captura de resultados	◆ Todas las coordinaciones estatales de evaluación ◆ Personal contratado	1. Documentación 2. Distribución de cuadernillos en las entidades	36. Manual técnico para la administración de cuadernillos y cuestionarios de contexto 37. Paquetes de cuadernillos distribuidos en las entidades federativas 38. Base de datos con resultados de la evaluación capturados
VI. Análisis e interpretación de los resultados de los Excale	13. Análisis iniciales de resultados	◆ Asesores en medición	1. Seminarios 2. Análisis estadísticos de resultados	39. Manual técnico sobre análisis de reactivos 40. Informe técnico sobre comportamiento psicométrico de reactivos
	14. Establecimiento de niveles de logro	◆ Comités de niveles de logro (uno por prueba)	1. Documentación 2. Capacitación 3. Trabajo colegiado	41. Manual técnico sobre establecimiento de niveles de logro 42. Documento donde se consignen los niveles de logro y puntos de corte de cada prueba
	15. Elaboración de informe técnico de resultados de aprendizaje	◆ Asesores en medición y validación	1. Documentación 2. Capacitación 3. Trabajo colegiado	43. Manual técnico sobre informe de resultados de aprendizaje 44. Informe técnico sobre resultados de aprendizaje asociados con variables de contexto
VII. Validación de los Excale	16. Estudios de validez de procesos y resultados de las pruebas	◆ Asesores en validación	1. Documentación 2. Investigación	45. Marco de referencia de los estudios de validez 46. Marco de referencia de cada prueba 47. Reportes técnicos de estudios de validez 48. Publicaciones sobre la validez de las pruebas

A continuación se describe brevemente cada una de las siete fases y 16 etapas del proceso de diseño, construcción, aplicación y validación de las pruebas Excale.

FASE I: Planeación general. La primera fase del proceso tiene el propósito principal de establecer un plan de pruebas a largo plazo. En esta fase es muy importante destacar la participación del *Consejo Técnico*⁴ y de los asesores externos en medición y validación, ya que en esta fase se definen los propósitos, principios y referentes conceptuales de las

⁴ Órgano colegiado constituido por 16 especialistas, nacionales y extranjeros, de gran prestigio por sus aportaciones a los campos de la evaluación y la investigación educativas.

pruebas, de los cuales dependen las fases restantes del proceso. Este apartado consta de tres etapas. En la primera se diseña el Plan General de Evaluación del Aprendizaje, donde se plasman el marco de referencia de los Excale, el proceso de diseño, construcción, aplicación y validación de pruebas, el diseño matricial y el modelo de niveles de logro; asimismo, se elabora el marco teórico y el programa para validar las interpretaciones de los Excale (INEE, 2004b). En la segunda etapa se diseñan y elaboran los cuestionarios de contexto dirigidos a estudiantes, docentes y directores de escuela, que se aplican junto con las pruebas para poder explicar los resultados de aprendizaje. Como se señaló anteriormente, las variables de contexto que exploran los cuestionarios se fundamentan en el modelo de aprendizaje escolar que adoptó el INEE; de particular relevancia serán aquellas relacionadas con las *oportunidades de aprendizaje*, que en la figura 1 tienen que ver con el currículum implementado. Finalmente, en la etapa 3 se diseña y desarrolla un sistema de base de datos para alojar, mantener y manejar la información relativa a las diferentes pruebas que se vayan generando, tales como: la estructura curricular de la asignatura, la tabla de contenidos de la prueba, las especificaciones y plantillas de reactivos, los resultados de la validez y los parámetros del piloteo de reactivos, entre otros.

FASE II: Estructuración de los Excale. A partir de la segunda fase, todas las etapas del proceso son específicas para cada prueba. Esta fase tiene el propósito de diseñar y justificar la estructura del examen (en una tabla de contenidos) y de elaborar las especificaciones de todos los reactivos que conformarán la prueba. En ella participan especialistas en currículum, en la enseñanza de la disciplina, autores de libros de texto, representantes de asociaciones, y docentes en ejercicio de distintos estratos y modalidades educativos⁵. En la cuarta etapa, el Comité Académico de cada prueba, formado por aproximadamente una decena de especialistas, realiza un análisis curricular exhaustivo de la asignatura y grado correspondientes, a partir del cual se elabora una retícula⁶ de la asignatura donde se explicita la estructura de los contenidos curriculares, así como los contenidos que son esenciales y muy importantes para evaluar. A partir de la retícula, se elabora la tabla de contenidos de la prueba y la justificación de los contenidos curriculares que la conforman. En esta tabla se explicitan las áreas temáticas, los temas y subtemas que derivan en los contenidos curriculares y habilidades intelectuales a evaluar, así como la forma y número en que serán evaluados. En la etapa 5, el Comité de Elaboración de Especificaciones de Reactivos de cada prueba, formado por otra decena de especialistas y docentes, define y describe en forma

⁵ Los estratos educativos para primaria son: escuelas públicas urbanas, escuelas públicas rurales, cursos comunitarios, escuelas privadas y educación indígena. Las modalidades educativas de secundaria son: secundarias privadas, secundarias generales, secundarias técnicas y telesecundarias.

⁶ Tabla de doble entrada donde se consignan los contenidos de una asignatura y sus relaciones, de acuerdo a sus ejes temáticos y tiempos pedagógicos (véase Contreras, 2000; Robredo, Ledesma y Alvarado, 1983).

detallada las características que debe tener cada reactivo. Es decir, se precisa el contenido a evaluar, su ubicación en el currículum, su importancia, la habilidad intelectual requerida al alumno, así como el formato del reactivo mismo (características de forma, fondo y redacción, entre otras), de tal manera que esta descripción sirva como guía o molde para construir reactivos similares y, hasta donde sea posible, equivalentes. El número de especificaciones para cada prueba es variable, pues depende de la extensión curricular de la asignatura⁷.

FASE III: Construcción de reactivos de los Excale. La tercera fase del proceso corresponde a la elaboración de los reactivos, así como a su validación y piloteo. Su propósito es producir reactivos de alta calidad, que conformarán cada prueba. En la sexta etapa, los miembros del Comité Constructor de Reactivos de cada prueba –constituido aproximadamente por una decena de autores de libros de texto y docentes en ejercicio– formulan de manera independiente tres reactivos para cada especificación⁸, los cuales serán revisados por otros especialistas conforme al manual de revisión de reactivos, que hace énfasis principalmente en la congruencia reactivo-especificación. El proceso se repite hasta que el reactivo quede a satisfacción de los revisores. Los reactivos resultantes se preeditan, se alojan en el sistema informático y pasan al Comité de Validez y Sesgo, de la etapa 7, constituido por docentes en ejercicio de las 32 entidades federativas del país y representantes de los diversos estratos y modalidades educativos. Este comité revisa cada reactivo en relación con su contenido (pertinencia curricular, grado de dificultad, entre otros), redacción (lenguaje utilizado, construcción sintáctica, por ejemplo) y sesgos cultural y de género (características de contenido y redacción que no favorezcan o perjudiquen a ningún grupo social). Cada reactivo es revisado independientemente por dos profesores y, si existe discrepancia, por un grupo de ocho docentes quienes debe llegar a un consenso. En su caso, se proponen recomendaciones para mejorar los reactivos o se justifica su eliminación. En la octava etapa se realiza un estudio piloto con la muestra intencional de aproximadamente cinco mil estudiantes⁹ por prueba, con el fin de conocer el comportamiento psicométrico de los reactivos y detectar los problemas que enfrentan los alumnos al responderlos. Este estudio piloto simula las condiciones reales de aplicación, que incluye la capacitación de coordinadores y responsables de evaluación, la aplicación de encuestas a alumnos, docentes y directores, así como la logística completa de los estudios evaluativos. La etapa culmina con la lectura de las respuestas de los estudiantes y la conformación de una base de datos con los resultados obtenidos.

⁷ Las pruebas de Español y Matemáticas aplicadas en 2005 tienen aproximadamente 130 especificaciones cada una.

⁸ Para cada prueba se desarrollan aproximadamente 390 reactivos.

⁹ La muestra de 2005 abarca cinco estados de la República, así como las modalidades y estratos educativos de mayor relevancia.

FASE IV: Conformación de los Excale. El propósito de esta fase es editar los cuadernillos de cada una de las pruebas nacionales, con la información recabada en las dos etapas anteriores. En esta etapa se seleccionan los reactivos que conformarán las pruebas nacionales, los cuales se agrupan en bloques¹⁰ y posteriormente, se definen las diversas formas¹¹ de cada una de las pruebas Excale. El resultado final de la etapa 9 es la selección de los mejores reactivos y su agrupación en bloques y formas que constituirán, de esta manera, las pruebas nacionales. Para la selección de reactivos, se analiza su comportamiento psicométrico (en diversos grupos de estudiantes) y la opinión de docentes en cuanto a su validez de contenido y ausencia de sesgo. Este análisis se lleva a cabo considerando los estándares de las pruebas criteriosales; es decir, preponderando el contenido sobre el comportamiento psicométrico. En la décima etapa se forman diferentes bloques de reactivos y con ellos distintas formas. Los bloques se conforman con una pequeña cantidad de reactivos, que pueden ser contestados en 15 minutos, y que comparten ciertas características que los hacen equivalentes en cuanto a su contenido, nivel de dificultad y varianza, así como en su extensión. Una vez conformados los bloques de reactivos, se combinan con el fin de armar distintas formas de prueba, de tal manera que todos los bloques se encuentren, hasta donde sea posible, en igual proporción y distribución; las formas pueden variar de una prueba a otra.¹² Finalmente, a cada forma se le agrega el cuestionario de contexto del alumno para mandar a impresión los cuadernillos resultantes, dando por concluida esta fase.

FASE V: Aplicación de los Excale. La quinta fase del proceso tiene como propósito principal administrar las pruebas a una muestra nacional de estudiantes y los cuestionarios de contexto a estudiantes, docentes y directores, así como capturar los resultados en una base de datos para su análisis posterior. Durante esta fase, se diseña la muestra poblacional de estudiantes de acuerdo a los estudios evaluativos que se planeen realizar; se capacita al personal encargado de la coordinación de la aplicación en las entidades federativas; se distribuyen los cuadernillos y se aplican las pruebas nacionales a las muestras poblacionales seleccionadas. El diseño de estas muestras se realiza en la etapa 11, acción que implica la selección de escuelas y estudiantes provenientes de las 32 entidades federativas del país. Dicho diseño depende de muchos factores, entre los que se encuentran: el marco muestral disponible, los propósitos del(los) estudio(s) a realizarse, la representatividad de las subpoblaciones que se deseen evaluar (rurales, urbanas y privadas, entre otras), el grado de confianza de la muestra y el tipo de diseño matricial de las pruebas. Una vez diseñada la muestra aleatoria, se valida con la información que proporcionan las escuelas seleccio-

¹⁰ Un bloque de reactivos está constituido por un grupo de preguntas que comparten características en común.

¹¹ Una forma está compuesta por la combinación de bloques de reactivos de distintas áreas temáticas.

¹² En el caso de las pruebas de Español y Matemáticas de 6º de primaria se construyeron 28 formas; para las mismas pruebas de 3º de secundaria se construyeron 24 formas.

nadas¹³. En la doceava etapa se realiza el proceso de capacitación a los coordinadores y responsables de la aplicación; se distribuyen los cuadernillos a las 32 entidades federativas; se administran las pruebas a los estudiantes seleccionados; se aplican los cuestionarios de contexto a docentes y directores de escuelas de los dos niveles educativos¹⁴. Por último, se leen y capturan los resultados de las pruebas y cuestionarios en una base de datos. Para el caso de los reactivos de respuesta abierta, donde se utiliza sólo una muestra nacional¹⁵ sin representatividad estatal –como es el caso de la sección de redacción de la prueba de Español– dos jueces califican cada reactivo, con base en rúbricas definidas en un protocolo de calificación, para posteriormente agregar los resultados a la base de datos correspondiente. Dada la magnitud de la operación¹⁶, esta fase del proceso es muy delicada, pues podrían presentarse problemas que invalidarían los resultados de las evaluaciones.

FASE VI: Análisis e interpretación de resultados de los Excale. Esta fase consta de tres etapas y tiene como propósito final la elaboración de los informes técnicos sobre los resultados de pruebas y cuestionarios de contexto, los cuales serán el sustento de los informes anuales, relacionados con el aprendizaje, que publicará el INEE. No obstante, para lograr esta meta, primero debe realizarse un análisis inicial de los resultados, con el fin de poder establecer los niveles de logro de los estudiantes e incluirlos en los reportes técnicos correspondientes. Así, en la etapa 13 se analizan estadísticamente los datos de la aplicación nacional y se realizan los primeros análisis estadísticos tanto de estudiantes como de reactivos, poniendo especial énfasis en el comportamiento psicométrico de estos últimos. En la etapa 14, el Comité de Establecimiento de Niveles de Logro de cada prueba –compuesto por autoridades educativas, especialistas en currículum, autores de libros de texto y docentes en ejercicio– define las categorías y niveles de logro de los estudiantes de acuerdo a: 1) la ejecución que presuntamente se esperaría de ellos y 2) los resultados reales en la prueba respectiva. Con esta información se definen los puntos de corte (puntuaciones mínima y máxima) de cada categoría. Finalmente, en la etapa 15 se elaboran los informes técnicos sobre los resultados de aprendizaje de los estudiantes evaluados, considerando las variables de contexto y de oportunidades de aprendizaje, recabadas en los cuestionarios de contexto de alumnos, docentes y directores. Al INEE le interesa conocer especialmente el logro del aprendizaje en las distintas subpoblaciones muestreadas (entidades federativas y estratos y modalidades educativos), así como los contenidos que los estudiantes dominan

¹³ Para la aplicación nacional de 2005 se seleccionaron aproximadamente 2 mil 800 escuelas de 6º de primaria y 2 mil 400 de 3º de secundaria, evaluando a 110 mil alumnos aproximadamente.

¹⁴ Se encuestarán cerca de 26 mil docentes de nivel primaria y 14 mil de nivel secundaria, así como a unos 4 mil 900 directores de escuelas de los dos niveles educativos (con excepción de los cursos comunitarios).

¹⁵ Cinco mil estudiantes por nivel educativo, aproximadamente, para la aplicación de 2005.

¹⁶ En esta etapa intervienen aproximadamente 5 mil 200 personas encargadas de aplicar las pruebas y cuestionarios.

del currículum nacional. No está por demás subrayar que la información generada en esta fase nutre significativamente la publicación del informe anual del INEE sobre *La Calidad de la Educación en México*.

FASE VII: Validación de los Excale. La última fase consta de una sola etapa y tiene como meta aportar información diversa sobre la validez de las pruebas Excale y de las interpretaciones que se deriven de su uso. Si bien, por fines prácticos, la etapa 16 se ubica al final del proceso de construcción de las pruebas, en realidad inicia desde el momento en que termina el diseño de una prueba (etapa 4). Es decir, podemos ver dos momentos de los estudios de validez de los Excale: 1) durante el proceso mismo de su construcción y 2) posterior a su terminación. En el primer caso, los estudios van dirigidos a evaluar la calidad y congruencia de cada una de las etapas del proceso, y tiene como finalidad verificar dicho proceso y generar información tendiente a corregir los problemas detectados mientras se construye la prueba. En el segundo caso, los estudios tienen el propósito de aportar evidencias de la veracidad y limitaciones de las interpretaciones que se generen como resultado del uso de una prueba. Su finalidad es doble: por un lado, legitimar académicamente las evaluaciones generadas por el INEE y, por el otro, iniciar un proceso de mejora de las pruebas, con base en la información y documentación de estos estudios. En sí, el proceso de validación de una prueba nunca termina.

Plan de Evaluación del Aprendizaje

Considerando que la misión del INEE es evaluar al Sistema Educativo Nacional, que los resultados de aprendizaje no cambian significativamente de un año a otro, y que la aplicación de pruebas de gran escala a nivel nacional es muy onerosa, se diseñó un esquema de evaluación racional, con el cual se puede hacer un seguimiento de las generaciones de estudiantes a lo largo de su escolaridad.

El plan de evaluación de los Excale se basa en un programa cuatrianual; es decir, un mismo grado escolar se evaluará cada cuatro años. La instrumentación de este esquema será gradual, de manera que poco a poco se evaluarán a todos los grados y áreas temáticas, como se muestra en la tabla II.

Tabla II. Programa de Evaluación del Aprendizaje del INEE

Grados	Años escolares											
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
3° Preescolar			E, M				E, M				E, M	
3° Primaria		E, M N, S				E, M N, S				E, M N, S		
6° Primaria	E, M		E, M		E, M N, S		E, M		E, M N, S		E, M	
3° Secundaria	E, M			E, M N, S		E, M		E, M N, S		E, M		E, M N, S
3° Bachillerato				E, M			E, M N, S				E, M N, S	
Evaluaciones especiales	PEM PECL											

Nota: E = Excale-Español; M = Excale-Matemáticas; N = Excale-Ciencias Naturales; S = Excale-Ciencias Sociales; PEM = Pruebas de Estándares Nacionales de Matemáticas; PECL = Pruebas de Estándares Nacionales de Comprensión Lectora.

En la tabla II se contempla que en 2005 se aplicarán los Excale de Español y Matemáticas en los grados de 6° de primaria y 3° de secundaria. En 2006 se evaluarán, además, Ciencias Sociales y Ciencias Naturales en 3° de primaria. En 2007 se evaluarán las habilidades de Lenguaje y Pensamiento Matemático en 3° de preescolar. Posteriormente, en 2008 se evaluarán cuatro asignaturas en 3° de secundaria y Matemáticas y Español en el bachillerato. A partir de 2009 se aplicarán todas las áreas previstas en el modelo en forma regular.

Como se puede apreciar, esta propuesta deja abierta la posibilidad de que se apliquen otras evaluaciones de aprendizaje, tales como alguna prueba de Comprensión Lectora y otra de Habilidad Matemática en algunos grados y modalidades educativas que el Consejo Técnico crea conveniente (por ejemplo, las Pruebas de Estándares Nacionales de 6° de primaria y 3° de secundaria para el estudio comparativo 2000 - 2005).

Conclusiones

Queda claro que para el INEE el objetivo fundamental de la evaluación del aprendizaje es conocer el rendimiento académico de los estudiantes a niveles estatal y nacional, así como los factores de contexto que lo explican. Lo anterior, con el fin de emitir juicios de valor que apoyen la toma de decisiones documentada y contribuir a realizar una rendición de cuentas a la sociedad mexicana sobre el estado que guarda la educación nacional (Poder Ejecutivo/ 2002).

Asimismo, se espera que los resultados de las evaluaciones que realice el INEE impacten los siguientes ámbitos educativos: políticas nacionales y estatales, currículum nacional,

programas de estudio y libros de texto, administración y gestión escolares, formación y actualización docentes, y opinión de la sociedad (INEE, 2004b). Para lograr este impacto, el INEE debe contar con un modelo del aprendizaje escolar que se adapte a las necesidades nacionales, así como con instrumentos y procedimientos técnica y metodológicamente sólidos.

Es importante subrayar dos características de la metodología empleada para diseñar, construir y validar las pruebas Excale: 1) el trabajo colegiado, donde intervienen una gran cantidad de especialistas y docentes en ejercicio, y 2) la documentación que se genera a lo largo de todo del proceso, la cual deja evidencia de su validez.

Respecto al trabajo colegiado, hay que destacar la participación de cinco comités, cada uno integrado aproximadamente por diez personas; la asesoría de dos grupos de expertos, conformado cada uno por tres especialistas; el trabajo por encargo llevado a cabo generalmente por cinco profesionales, y el trabajo especializado de cerca de 25 técnicos de la Dirección de Pruebas y Medición. Es decir, para el diseño, construcción y validación de cada prueba Excale se requiere de casi un centenar de especialistas y docentes, aparte de las aproximadamente 900 personas que intervienen en el estudio piloto y las cerca de 5,200 que interviene en la aplicación nacional.

Respecto a la documentación que se genera con esta metodología, debemos destacar tres aspectos. Primero, en su mayoría, los productos se van elaborando en forma secuenciada y cada uno de ellos sirve de insumo para el siguiente paso. Segundo, los productos generados en cada etapa son revisados y analizados por el siguiente comité, y aunque éste no puede modificar lo hecho por el grupo anterior, sí debe señalar y, en su caso, justificar las observaciones que se hagan al respecto para tener evidencia de los problemas que se deben resolver en el futuro para mejorar el instrumento. Tercero, la documentación producida al final del proceso es abundante y rica en información; cerca de 48 productos se elaboran para cada prueba, de los cuales aproximadamente la mitad son comunes a todas ellas (como los manuales técnicos) y la mitad restante son específicos de cada examen (como la tabla de contenidos de la prueba).

Con esta metodología se considera posible alcanzar dos de las grandes metas que se trazó el INEE respecto a la evaluación del aprendizaje:

1. Conocer el logro académico de los estudiantes a niveles estatal y nacional, así como los factores de contexto más importantes que lo determinan.
2. Contribuir a conocer los alcances y limitaciones del Sistema Educativo Nacional y, con ello, a elevar el nivel de la calidad de la educación básica México.

Para finalizar, debemos subrayar que este documento explica de manera general el modelo de las pruebas Excale, y representa un primer paso para dar a conocer al público los

aspectos más relevantes de la metodología empleada. Sin embargo, se deberá tomar con las reservas necesarias, pues, el modelo de evaluación aquí expuesto se irá modificando y fortaleciendo conforme se reciba información de los resultados de su instrumentación.

El Instituto Nacional para la Evaluación de la Educación, siendo coherente con sus principios, tiene el compromiso de informar a la sociedad de sus métodos y resultados, razón por la cual se publica este escrito. Lo anterior cobra una gran relevancia si consideramos que en el pasado las evaluaciones nacionales fueron consideradas *cajas negras*, cuyo acceso era muy restringido.

REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME). (1999). *Standards for Educational and Psychological Testing*. Washington: American Psychological Association.
- Backhoff, E. (2005). La comparación entre entidades: alcances y limitaciones de los *rankings*. En: *Memorias de las Jornadas de Evaluación Educativa*. México: Instituto Nacional para la Evaluación de la Educación (INEE).
- Backhoff, E., Andrade, E., Peón, M., Sánchez, A., Juárez, E., Monroy, L., y Tanamachi, M. (2003). *Resultados de las pruebas nacionales: comprensión lectora y matemáticas, 6° de primaria y 3° de secundaria*. México: Instituto Nacional para la Evaluación de la Educación (INEE).
- Backhoff E. y Martínez-Rizo, F. (2004). Resultados de las Pruebas de Estándares Nacionales 2003: elementos para la comparación entre entidades. En: *Memoria 2004, Sexto Foro de Evaluación Educativa*. México: Centro Nacional para la Evaluación de la Educación Superior (Ceneval).
- Backhoff, E., Monroy, L. y Tanamachi, L. (2005). *Cuestionarios de Contexto de las pruebas Excale: Marco de referencia*. Documento mimeografiado. INEE: México.
- Carroll, J. (1963). A model of school learning, *Teacher College Record*, 64, 723-733.
- Contreras, L. A. (2000). *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California*. Tesis para obtener el grado de Maestro en Ciencias Educativas. México: Universidad Autónoma de Baja California.
- Contreras, L. A., Backhoff, E. y Larrazolo, N. (2003). *Curso taller para la elaboración de exámenes criterios: manual para el Comité Diseñador del Examen*. Documento mimeografiado. Ensenada: Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California.

- Deng, H. Ferris, J. & Hombo, C. (2003). *A vertical Scheme of Building the NAEP Booklets*. Documento presentado en la Reunión Anual del National Council on Measurement for Education (NCME), Chicago.
- Dirección General de Evaluación (DGE). (2004). *La evaluación en la Secretaría de Educación Pública*. Documento Mimeografiado. México: SEP-DGE.
- Dirección General de Evaluación (DGE). (2000). *Balance de las acciones emprendidas entre diciembre de 1994 y octubre de 2000*. Documento recuperado el 24/05/2005 en: <http://www.sep.gob.mx/work/appsite/dge/index.htm>
- Gaviria, J. L. (2005). *Propuesta de diseño matricial para las pruebas de español y matemáticas del programa de pruebas nacionales del INEE*. Documento mimeografiado. México: Instituto Nacional para la Evaluación de la Educación (INEE).
- Instituto Nacional para la Evaluación de la Educación (INEE). (2004a). *La calidad de la educación básica en México: resultados de evaluación educativa 2004*. México: INEE.
- Instituto Nacional para la Evaluación de la Educación (INEE). (2004b). *Plan General de Evaluación del Aprendizaje*. Documento mimeografiado. México: INEE.
- Instituto Nacional para la Evaluación de la Educación (INEE). (2003). *La Calidad de la Educación Básica en México*. México: INEE.
- Martínez-Rizo, F. (2004). Comparabilidad de los resultados de las evaluaciones. En: *Memorias de las Jornadas de Evaluación Educativa*. México: Instituto Nacional para la Evaluación de la Educación (INEE).
- Martínez-Rizo, F., Backhoff, E., Castañeda, S., De la Orden, A., Schmelkes, S., Solano-Flores, G., Tristán, A. y Vidal, R. (2000). *Estándares de calidad para Instrumentos de Evaluación Educativa*. México: Centro Nacional para la Evaluación de la Educación Superior (Ceneval).
- Mullins, I., Martin, M., Smith, T., Garden, R., Gregory, K., González, E., Chrostowski, S. & O'Connor, K. (2003). *TIMSS, Assessment Frameworks and Specifications. 2003 (2a. ed.)*. Boston: Boston College.
- Nitko, A. (1994, julio). *A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students*. Po-

nencia presentada en la Conferencia Internacional sobre Evaluación y Medición Educativas, de la Asociación para el Estudio de la Evaluación Educativa en Sudáfrica (ASSESA).

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2003). *The PISA 2003, Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Documento recuperado el 24/05/2005 en: <http://www.pisa.oecd.org/dataoecd/46/14/33694881.pdf>

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2004). *Contextual Framework for PISA 2006*. Draft version. National Project Managers Meeting, Bratislava, Slovakia.

Poder Ejecutivo. (2002). Decreto de creación del Instituto Nacional para la Evaluación de la Educación. *Diario Oficial*, 08/08/2002, México, D.F.

Popham, J. (1990). *Modern Educational Measurement: A Practitioner's Perspective*. Englewood Cliffs, N. J.: Prentice-Hall.

Robredo, J. M., Ledesma, R. y Alvarado, F. (1983). *Reticulación: una estrategia para la elaboración de programas de estudio*. Tesis para obtener el grado de Licenciado en Psicología. México: Universidad Nacional Autónoma de México (UNAM).

Scheerens, J. (1997). *The foundations of School Effectiveness*. Oxford: Pergamon.

Secretaría de Educación Pública (SEP). (2001). *Programa Nacional de Educación 2001-2006*. México: SEP.

Van der Linden, W., Veldkamp, B. & Carlson, J. (2004). Optimizing Balanced Incomplete Block Designs for Educational Assessments. *Applied Psychological Measurement*, 28(5), 317-331.

Velázquez, V. (2000). Hacia una cultura de la evaluación. En SEP: *Memorias del quehacer educativo 1995-2000*. México: SEP.

Wright, B. & Stone, M. (1998). *Diseño de mejores pruebas: utilizando la técnica de Rasch*. México: Centro Nacional para la Evaluación de la Educación Superior (Ceneval).

ANEXO: PROGRAMAS EVALUATIVOS DE LA EDUCACIÓN BÁSICA A CARGO DE LA DGE

Programa	Propósito	Población meta	Instrumento	Inicio
Instrumento de Diagnóstico para Estudiantes de Nuevo Ingreso a la Secundaria (IDANIS)	Seleccionar y diagnosticar aspirantes que buscan su ingreso a la secundaria	Estudiantes que terminaron el 6° de primaria	Examen de habilidad verbal, matemática y de razonamiento formal	1989
Evaluación del factor de aprovechamiento escolar de los estudiantes	Evaluar a docentes inscritos en el Programa de Carrera Magisterial y a sus estudiantes	Muestras de estudiantes de 3° a 6° de primaria y de 1° a 3° de secundaria	Exámenes de contenidos programáticos	1994
Evaluación de la Educación Primaria (EVEP)	Evaluar los resultados del Programa para abatir el rezago de la Educación Básica (PAREB)	Estudiantes de 2° a 6° de primaria y de 1° a 3° de secundaria	Exámenes que antecedieron a las Pruebas de Estándares Nacionales	1996
Olimpiada del Conocimiento infantil	Darle reconocimiento a los alumnos sobresalientes	Estudiantes de 6° de primaria	Exámenes de distintas materias del currículum nacional	1997
Estándares Nacionales		Estudiantes de 2° a 6° de primaria y de 1° a 3° de secundaria	Examen de habilidades verbales y matemáticas	1998
Programa de Escuelas de Calidad (PEC)	Medir el nivel de logro de las habilidades de Comprensión Lectora y Matemáticas	Estudiantes del PEC de 3° a 6° de primaria y de 1° a 3° de secundaria	Examen de habilidades verbales y matemáticas	2002
Consejo Nacional de Fomento Educativo (CONAFE)		Estudiantes de 3° a 6° de primaria	Examen de habilidades verbales y matemáticas	2002
Diagnóstico censal	Conocer el nivel de desarrollo de conocimientos y habilidades	Estudiantes de todos los grados de primaria	Exámenes de conocimientos y habilidades de Español y Matemáticas	2004

Fuentes: Velázquez (2000) *Hacia una cultura de la evaluación 1994-2000*; DGE (2000), *Balance de las acciones emprendidas entre diciembre de 1994 y octubre de 2000*. Martínez-Dustan (2005), Comunicación personal.

TÍTULOS DE ESTA COLECCIÓN:

- 1.- *Los resultados de las pruebas de PISA*
Martínez Rizo, Felipe.
- 2.- *Factores externos e internos a las escuelas que influyen en el logro académico de los estudiantes de nivel primaria en México, 1998-2002*
Muñoz I, Carlos et al.
- 3.- *Contextualización sociocultural de las escuelas de la muestra de estándares nacionales (1998-2002).*
_____ *Determinantes sociales y organizacionales del aprendizaje en la Educación Primaria de México. Un análisis de tres niveles.*
_____ *Perfil de las escuelas primarias eficaces de México (2001)*
Fernández, Tabaré.
- 4.- *Tercer Estudio Internacional de Matemáticas y Ciencias Naturales (TIMSS): resultados de México en 1995 y 2000*
Backhoff, Eduardo y G. Solano.
- 5.- *Estudio Sobre las Desigualdades Educativas en México: la Incidencia de la Escuela en el Desempeño Académico de los Alumnos y el rol de los Docentes*
Treviño, Ernesto y G. Treviño.
- 6.- *Evaluación inicial de los procesos de calibración y equiparación de las pruebas del proyecto de estándares nacionales.*
Magriñá, Antonio.
- 7.- *Factores asociados al aprendizaje del lenguaje y las matemáticas en 13 estados de México.*
Cervini, Rubén.
- 8.- *Acciones de Evaluación en las Instituciones Públicas de Educación Media Superior*
Antonio, Rocío.
- 9.- *El proyecto PISA: su aplicación en México*
Vidal, Rafael, et. al.
- 10.- *La Comparabilidad de los Resultados de las Evaluaciones. Importancia y Dificultad de la Equiparación*
Martínez Rizo, Felipe.
- 11.- *Marginación y rezago educativo en México.*
Ávila, José Luis.
- 12.- *Pruebas y rendición de cuentas*
Martínez Rizo, Felipe.
- 13.- *Panorama Educativo 2004. La edición 2004 de Education at a Glance de la OCDE*
Martínez Rizo, Felipe.
- 14.- *El Diseño de Sistemas de Indicadores Educativos: Consideraciones Teórico- Metodológicas*
Martínez Rizo, Felipe.
- 15.- *Sobre la difusión de resultados por escuela*
Martínez Rizo, Felipe.
- 16.- *La telesecundaria mexicana: desarrollo y problemática actual*
Martínez Rizo, Felipe.
- 17.- *Exámenes de la Calidad y el Logro Educativos (Excale): nueva generación de pruebas nacionales*
Backhoff Escudero, Eduardo.
- 18.- *La educación mexicana en Education at a glance 2005*
Martínez Rizo, Felipe.