



INEE

Instituto Nacional para la
Evaluación de la Educación

Estudio comparativo del aprendizaje en tercero de secundaria en México

2005-2008

Español y Matemáticas



Resultados de
evaluaciones

Logro escolar

Aplicaciones de la prueba Excale



INEE

Instituto Nacional para la
Evaluación de la Educación

Estudio comparativo del aprendizaje en tercero de secundaria en México

2005-2008

Español y Matemáticas

*Estudio comparativo del aprendizaje
en tercero de secundaria en México 2005-2008. Español y Matemáticas*

Primera edición, 2012

ISBN: 978-607-7675-41-9

Instituto Nacional para la Evaluación de la Educación

D. R. © Instituto Nacional para la Evaluación de la Educación
José Ma. Velasco 101, Col. San José Insurgentes,
Delegación Benito Juárez, C.P. 03900, México, D. F.

COORDINADORES

Andrés Sánchez Moguel

Edgar Andrade Muñoz

COLABORADORES

Enrique Estrada Cruz

Marisela García Pacheco

Glenda Patricia Guevara Hernández

José Gustavo Rodríguez Jiménez

COORDINACIÓN EDITORIAL

María Norma Orduña Chávez

CORRECCIÓN DE ESTILO Y DISEÑO

Publicaciones Corporativas

CORRECTOR DE PRUEBAS

Juan Carlos Rosas Ramírez

Impreso y hecho en México.

Distribución gratuita. Prohibida su venta.

Consulte el catálogo de publicaciones en línea: www.inee.edu.mx

La elaboración de esta publicación estuvo a cargo de la Dirección de Pruebas y Medición. El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del editor. Se autoriza su reproducción parcial o total por cualquier sistema mecánico o electrónico para fines no comerciales y citando la fuente de la siguiente manera:

INEE (2012). *Estudio comparativo del aprendizaje en tercero de secundaria en México 2005-2008. Español y Matemáticas*. México: INEE.

Índice

5 Presentación

7 Introducción

- 7 Sentido y propósito de los Excale
- 8 Características generales de los Excale
- 9 Plan general de evaluación
- 10 Diseño, construcción, aplicación y análisis de los Excale
- 10 Sobre el diseño y la construcción
- 10 Sobre la aplicación de los Excale
- 12 Sobre el análisis de los resultados
- 12 Niveles de logro
- 13 Antecedentes y estructura del informe

Capítulo 1.

15 Procedimiento de equiparación

- 15 La comparabilidad de resultados 2005-2008
- 16 La enseñanza que dejó la anomalía de las pruebas de lectura de NAEP 1985-1986
- 17 Una experiencia más reciente de NAEP. Las pruebas de Ciencias de 2005
- 17 El diseño matricial de las pruebas Excale-09: 2005 y 2008
- 21 La comparación de Español y Matemáticas, 2005 y 2008

Capítulo 2.

24 Resultados y comparación en Español

- 24 Definición de la escala de Español
- 25 Resultados comparativos en Español
- 28 Comparación por modalidad
- 30 Brechas entre modalidades
- 30 Dispersión dentro de las modalidades
- 31 Comparación por sexo
- 32 Comparación por sexo y modalidad
- 32 Comparación por edad
- 33 Comparación por edad y modalidad

Capítulo 3.

36 Resultados y comparación en Matemáticas

- 36 Definición de la escala de Matemáticas
- 36 Rediseño del Excale de Matemáticas
- 37 Resultados comparativos en Matemáticas
- 41 Comparación por modalidad
- 42 Brechas entre modalidades
- 43 Dispersión dentro de las modalidades
- 44 Comparación por sexo
- 45 Comparación por sexo y modalidad
- 45 Comparación por edad
- 47 Comparación por edad y modalidad

Capítulo 4.

48 Desagregación por entidades

Capítulo 5.

51 Conclusiones

- 52 En Español
- 54 En Matemáticas
- 55 Sobre la información por entidad
- 55 Para terminar

56 Bibliografía

57 Índice de tablas

58 Índice de gráficas

59 Anexos

- 59 Anexo 1
Tablas complementarias de Español
- 61 Anexo 2
Tablas complementarias de Matemáticas

Presentación

Las pruebas Excale que el INEE aplica desde 2005 tienen un diseño y siguen un plan de trabajo que permite su comparación en el tiempo. Cada cuatro años se repite la evaluación de un determinado grado escolar y se lleva a cabo un trabajo metodológico específico que coloca las diferentes aplicaciones en una misma métrica para mostrar si en ese tiempo el nivel de logro de los estudiantes ha mejorado, empeorado o permanece igual.

Este informe presenta un análisis comparativo de las aplicaciones 2005 y 2008 de las pruebas Excale de tercero de secundaria. Informes anteriores dan cuenta de cada una de estas dos aplicaciones por separado, de manera que este volumen incluye los resultados solamente desde una perspectiva dinámica. No es posible hablar aún de tendencias, pues esto requeriría más puntos en el tiempo, sin embargo, consideramos importante continuar el esfuerzo de presentar los cambios en el logro a nivel nacional que el INEE ha realizado con la publicación *Estudio Comparativo de la Educación Básica en México: 2000-2005*, utilizando las Pruebas de Estándares Nacionales elaboradas por la entonces Dirección General de Evaluación de la SEP, y del *Estudio Comparativo del Aprendizaje en Sexto de Primaria en México 2005-2007* publicado por el INEE en 2008 que ya utilizó las pruebas Excale.

Además de la comparación en sí, este volumen presenta en el capítulo 1, de manera detallada, el método que se utiliza para comparar las pruebas Excale de diferentes años y del mismo grado escolar. Se hizo un esfuerzo por lograr que este capítulo tuviera el mínimo de contenido técnico posible, de modo que fuera accesible a muchos ti-

pos de lectores, en el interés de que más personas conozcan con claridad la manera en que se hacen estas comparaciones y puedan juzgar si éstas son pertinentes o no. En posteriores publicaciones haremos referencia a lo explicitado aquí mientras no cambie el procedimiento de comparación.

Los capítulos 2 y 3 muestran los resultados comparativos de los dos años estudiados en Español y Matemáticas, respectivamente. Aunque en 2008 se realizaron también evaluaciones de Biología y Formación cívica y ética, las primeras son las asignaturas que se evaluaron en 2005 y, por ello, la comparación con 2008 sólo es posible en estos dos casos.

Las comparaciones en el tiempo se desagregan por modalidad, sexo y edad en los capítulos 2 y 3. El capítulo 4 presenta algunos datos desagregados por entidad en las dos aplicaciones para complementar la información.

Finalmente, el capítulo 5 presenta las conclusiones. Si bien el lector podrá encontrar una serie de elementos para la reflexión, la conclusión fundamental del estudio es que en general no se observan cambios en los niveles de logro entre 2005 y

2008. Este resultado debe ser matizado al considerar el aumento en la matrícula en este periodo, ya que esto significa que ingresaron a la escuela secundaria educandos que son los primeros en sus familias en cursar este nivel educativo.

No obstante los argumentos que sea plausible construir para explicar que no haya habido cambios en los niveles de logro, el Sistema Educativo Nacional es responsable de diseñar acciones de política enfocadas a la mejora de los aprendizajes, pues tanto las pruebas nacionales como Excale o ENLACE, y las internacionales como PISA, han mostrado a lo largo del tiempo que proporciones muy importantes de estudiantes no están alcanzando los niveles apropiados de logro, y que los cambios a este respecto han sido lentos o inexistentes.

¿Cuánto tiempo requiere un sistema educativo tan grande como el mexicano para que la mayoría de sus estudiantes consigan niveles de logro que les permitan continuar aprendiendo a lo largo de su vida? ¿Qué acciones a nivel nacional, estatal y escolar son necesarias para lograrlo? Éstas y otras reflexiones se derivan de los resultados que arroja el presente estudio; el INEE confía en que este documento nutra la discusión y contribuya a la construcción de propuestas orientadas a la mejora de los resultados de aprendizaje de todos nuestros estudiantes.

Margarita Zorrilla Fierro

Directora General



Introducción

El propósito de este volumen es presentar una comparación entre los resultados de los Exámenes para la Calidad y el Logro Educativos aplicados en 2005 y en 2008 a estudiantes de 3° de secundaria (se llamará a dichos exámenes Excale 09; el número indica el grado escolar evaluado). Estas dos aplicaciones son las únicas que se han realizado de las pruebas Excale para dicho grado escolar. Además, las pruebas de 2005 se hicieron sólo para las asignaturas de Español y Matemáticas, mientras que en el caso de 2008 se incluyeron las asignaturas de Biología y Formación cívica y ética, por ello, la comparación sólo es posible para las dos primeras.

Un propósito ligado al anterior es presentar con cierto nivel de detalle el procedimiento empleado hasta ahora para hacer comparables en el tiempo las pruebas Excale, de modo que este documento sirva como referente del método empleado a este respecto.

Este reporte complementa los resultados a detalle de la aplicación 2008, presentados en el documento *El aprendizaje en tercero de secundaria en México*. Informe sobre los resultados de Excale 09, aplicación 2008 (Sánchez y Andrade, 2009), que se encuentra disponible en la página electrónica del Instituto Nacional para la Evaluación de la Educación (INEE).

Desde el punto de vista de resultados a nivel nacional, el intervalo de tiempo analizado es pequeño, por lo tanto, no se esperan grandes cambios. Conforme se agreguen las aplicaciones de las pruebas Excale los datos se irán agregando para mostrar tendencias.

Sentido y propósito de los Excale¹

En el marco de la misión del INEE, al contribuir “al mejoramiento de la educación en México a través de la realización de evaluaciones integrales de la calidad del sistema educativo y de los factores que la determinan” (INEE, 2006: 23), los Excale tienen como propósito conocer lo que los estudiantes mexicanos aprenden del currículo a lo largo de su educación básica. A partir de dicha evaluación, el INEE aporta elementos para la reflexión, la toma de decisiones de política educativa y la rendición de cuentas a la sociedad mexicana.

El INEE reconoce que los profesores pueden hacer una evaluación a profundidad de los conoci-

¹ Se considera que la información presentada aquí es suficiente para la contextualización de resultados, sin embargo, una versión más detallada sobre las pruebas Excale, su sentido, propósito y características se encuentra en *El aprendizaje en tercero de secundaria en México*. Informe sobre los resultados de Excale 09, aplicación 2008 (Sánchez y Andrade, 2009).

mientos, habilidades y actitudes que desarrollan sus alumnos. Los Excale no están diseñados para hacer una evaluación individual de los estudiantes ni de las escuelas; su sentido es mostrar los resultados educativos a nivel de sistema, es decir, con respecto a grandes grupos de estudiantes: por modalidad, por entidad, por sexo, por edad. Además, estas pruebas permiten reconocer los contenidos educativos que los estudiantes, como conjunto, dominan y los que no dominan. Los Excale se han diseñado con la intención de hacer comparaciones en el tiempo. Las subsecuentes aplicaciones de las pruebas permitirán observar cambios en los contextos y en los aprendizajes de la población de alumnos de los grados evaluados, así como en las relaciones entre estos elementos.

Una buena práctica de la evaluación educativa, reconocida internacionalmente, es “especificar los usos apropiados de los resultados y prevenir a los usuarios de la información de posibles usos incorrectos”, tal como señala el Joint Committee on Testing Practices (JCTP, 2004: 8). Los resultados presentados en este reporte pueden ser utilizados de manera específica para analizar cambios en el tiempo en el logro educativo a nivel nacional y por las modalidades reportadas. En el transcurso del tiempo, los cambios detectados pueden mostrar tendencias que esclarezcan el desarrollo de la educación básica a nivel nacional. También es posible comparar los aprendizajes de hombres y mujeres con el propósito de monitorear posibles inequidades, así como los aprendizajes de alumnos en edad normativa y con extraedad como un insumo para la reflexión sobre procesos (por ejemplo, la reprobación y el ingreso tardío a la educación básica).

Debido a los propósitos, diseño, construcción y aplicación de los Excale, sus resultados no deben ser utilizados para:

- Evaluar individualmente a los estudiantes, ni para evaluarlos como grupo escolar de una misma aula, como grupo de estudiantes de una misma escuela o de una misma zona escolar.
- Evaluar indirectamente el desempeño de profesores o escuelas a partir del logro de los alumnos.
- Evaluar el efecto de programas de mejora educativa estatales o federales.

- Hacer comparaciones entre diferentes asignaturas.

Para llevar a cabo correctamente su tarea, la Dirección de Pruebas y Medición del INEE desarrolló un plan de trabajo que indica los años de aplicación de las diferentes pruebas, además de establecer las características y el método de trabajo para garantizar la calidad de los Excale. A continuación se presentan de manera sucinta estos elementos.

Características generales de los Excale

Los Excale son pruebas criterioles, alineadas al currículo nacional, cuya aplicación se basa en un arreglo matricial. Son criterioles porque tienen un referente fijo, establecido antes de aplicar las pruebas. Dicho referente es el currículo nacional. El que las pruebas estén alineadas al currículo significa que sólo incluyen contenidos que se encuentran en el mismo —la mayor cantidad posible de éstos—. Los equipos que diseñan y desarrollan las pruebas Excale cuidan que los contenidos curriculares estén bien representados en las preguntas utilizadas, tanto en la temática y la extensión como en el enfoque, que deben ser los que la Secretaría de Educación Pública (SEP) explicita en materiales curriculares como planes y programas, libros de texto de los alumnos, libros del profesor, ficheros y talleres nacionales, entre otros.

Cuando se desea evaluar extensamente los contenidos curriculares, es común desarrollar pruebas que tienen una gran cantidad de ítems. Si el número de preguntas sobrepasa la cantidad a la que puede ser expuesto un niño sin perder atención o sentir tedio por la tarea, se corre el riesgo de tener resultados sesgados por el cansancio. Ante ello, una posibilidad sería la aplicación de las pruebas en sesiones cortas, a lo largo de varios días, lo cual tendría inconvenientes, entre los que se destacan, por ejemplo, fuerte aumento en el presupuesto requerido para la aplicación y la dificultad de asegurar la integridad de la muestra, es decir, que los niños participantes en el primer día de aplicación de las pruebas continúen con estas en días sucesivos, hasta terminarlas. Puesto que los Excale no tienen el propósito de dar resultados por alumno, ni por escuela, existe una segunda manera de desprenderse de la necesidad de exponer a los niños a largas sesiones de evaluación: utilizar un arreglo matricial.

Plan de Evaluación del logro con Excale 2004-2016

Grados	Años escolares												
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
3º Preescolar				E,M				E,M				E,M	
3º Primaria			E,M,N,S				E,M,N,S				E,M,N,S		
6º Primaria		E,M		E,M		E,M,N,S				E,M,N,S			
3º Secundaria		E,M			E,M,N,S				E,M,N,S				E,M,N,S
Último grado de Media Superior							E,S						

Notas

E= Excale de Español
M= Excale de Matemáticas
N= Excale de Ciencias Naturales
S= Excale de Ciencias Sociales

Proceso del Excale



El arreglo matricial de los reactivos de las pruebas Excale consiste en repartir las preguntas de una misma escala en diferentes cuadernillos. Así, cada estudiante es expuesto a una cantidad limitada de ítems, adecuada a su edad y grado escolar. Los diferentes cuadernillos tienen diferentes grupos de preguntas, de tal manera que cada pregunta está en varios cuadernillos. Se atiende la variedad temática y el equilibrio en la dificultad de las distintas preguntas de un mismo cuadernillo. Por ende, es posible utilizar las preguntas contestadas por los estudiantes para establecer una sola escala en la cual todos los reactivos se relacionan entre sí y conocer cuáles son los contenidos que los alumnos dominan como conjunto y cuáles no.

Plan general de evaluación

Actualmente, los Excale evalúan cuatro grandes áreas curriculares: Español, Matemáticas, Ciencias naturales y Ciencias sociales. Solamente se evalúan los grados terminales de cada nivel (tercero de preescolar, sexto de primaria, tercero de secundaria) y también el tercero de primaria, para balancear el seguimiento de los aprendizajes en periodos de tres años. Es importante comentar que, aunque la evaluación se hace al finalizar los grados señalados, los contenidos curriculares considerados para el diseño de las pruebas incluyen los dos grados previos.

Además, las pruebas Excale de los diferentes grados y asignaturas no se aplican cada año. Debido a que el propósito principal de los Excale es evaluar el sistema educativo en su conjunto, se considera que esta evaluación puede hacerse en ciclos de cuatro años; en efecto, debido a las inercias de todo un sistema educativo, es poco probable que grandes cambios se lleven a cabo año con año, siendo más factible que pequeños cambios puedan verse tras varios años. Con la intención de focalizar esfuerzos, estos ciclos de cuatro años se escalonan de modo tal que cada año se evalúa un grado escolar diferente.

La figura 1 muestra en qué años se llevan a cabo los diferentes procesos de cada prueba, incluyendo su aplicación. En esta figura, cada barra señala el ciclo completo de una prueba desde su construcción hasta el análisis y su reporte final. Un momento crucial es el del año en que se lleva a cabo la aplicación, que se encuentra en medio de cada ciclo. Si se mira una columna en particular, se puede observar que en cada año se está construyendo una prueba, aplicando otra y analizando otra más².

² El cuadro que aparece en el cruce de sexto de primaria y 2007 indica la réplica de los Excale de Español y Matemáticas diseñados en 2004, aplicados en 2005 y reaplicados en 2007. La intención era tener pronto datos que permitieran hacer comparaciones en el tiempo. Dado que es un evento que sale de la lógica general del plan de pruebas, no se ha marcado de la misma manera que los demás ciclos.

Las mismas asignaturas en el mismo grado son evaluadas cada cuatro años. Esto permite hacer series de tiempo que contrastan los resultados de muestras similares ante pruebas comparables entre sí, debido a que los Excale cuentan con un diseño que prevé su comparabilidad en aplicaciones sucesivas. Este análisis es importante para contestar preguntas como “¿mejora en el tiempo el logro educativo de los alumnos con respecto a los aprendizajes curriculares a nivel nacional?”.

Diseño, construcción, aplicación y análisis de los Excale

Existe una serie de documentos técnicos que explican de manera extensa y detallada las diferentes fases de trabajo con los Excale³. En este apartado se comentan algunos elementos cruciales del proceso para la comparación de resultados 2005-2008 de tercero de secundaria.

Sobre el diseño y la construcción

Los Excale 09 de Español y Matemáticas aplicados en 2008 son versiones rediseñadas de las pruebas aplicadas en 2005. Tras la primera aplicación y el reporte, se encontraron nuevas interrogantes y elementos que parecía deseable agregar a las pruebas. Por esta razón, durante 2007 y principios de 2008 se llevó a cabo el rediseño de las pruebas de estas dos asignaturas y, para ello, se consideró el conjunto de contenidos que se había utilizado originalmente; sin embargo, se eliminaron algunos que no mostraron utilidad, se reformularon algunos otros y se incluyeron algunos nuevos contenidos de acuerdo con las necesidades de cada asignatura. En todos los casos, los comités Académico, de Especificaciones y de Construcción de reactivos fueron conscientes de que la prueba tenía un antecedente y que la intención era mejorar la escala, de manera que no se perdiera la comparabilidad entre las dos aplicaciones para el correcto análisis de sus resultados como una serie de tiempo.

Es importante hacer notar que se decidió utilizar como base para el rediseño de estos dos Excale

³ Por ejemplo, Backhoff *et al.* (2005). *Exámenes de la calidad y el logro educativos (Excale): nueva generación de pruebas nacionales*. Colección cuadernos de investigación. Cuaderno no. 17. México: INEE; también deben citarse los manuales técnicos dedicados a cada proceso de construcción de las pruebas, que pueden consultarse en: http://www.inee.edu.mx/index.php?option=com_content&view=category&layout=blog&id=382&Itemid=1196

el currículo previo a la reforma curricular, es decir, el de 1993, en virtud de que la generación evaluada en 2008 es la última que se formó con los planes y programas de dicho año. Como se comentó anteriormente, una de las características de los Excale es que están alineados al currículo, por lo tanto es justo evaluar lo que el currículo señala que los estudiantes deben aprender a su paso por la secundaria. Sin embargo, dado que ya se conocía la propuesta de cambio curricular desde que inició el rediseño de los Excale 09 de Español y Matemáticas, se procuró enfatizar los contenidos que perduran en la reforma, con miras a permitir comparaciones con posteriores aplicaciones de los Excale 09.

Sobre la aplicación de los Excale

Al igual que otros estudios que realiza el INEE, los resultados presentados en este informe se infirieron a partir de las respuestas de muestras de alumnos. Desde sus inicios, el INEE tomó la decisión de realizar estudios del logro educativo utilizando muestras por las siguientes razones:

- Los estudios muestrales resultan, con un margen bastante amplio, menos onerosos que los censales.
- Se puede definir la precisión de los resultados obtenidos con una muestra, con el fin de evaluar al Sistema Educativo Nacional de forma adecuada.
- Se tiene mayor control sobre la administración de los instrumentos de evaluación, lo cual disminuye el riesgo de captar sesgos no cuantificables en detrimento de la calidad de la información a recolectar:
 - o En una muestra es factible capacitar adecuadamente a todo el personal involucrado en el manejo de los instrumentos. En un censo es fácil perder el control de quién está participando.
 - o En una muestra, las condiciones en que se resuelven las pruebas son más homogéneas que en los censos.
 - o En una muestra se minimiza el riesgo de la pérdida de material y la posterior divulgación inadecuada del mismo.

Los estudiantes que participaron en el estudio fueron seleccionados como muestras de los alumnos del tercer grado de secundaria del Sistema Edu-

cativo Nacional inscritos en los ciclos escolares 2004-2005 y 2007-2008. El número de escuelas y la matrícula en estos dos ciclos se pueden ver en la tabla I.

En 2008 hubo un incremento importante a nivel nacional (9%) en el número de escuelas, y en menor medida (4%) en la matrícula de tercero de secundaria. El número de escuelas de sostenimiento privado aumentó en mayor medida (14%), seguido por las secundarias generales (8%), técnicas (5%) y telesecundarias (3%). El número de escuelas secundarias para trabajadores disminuyó 4%.

A nivel nacional, el número de alumnos inscritos en tercero de secundaria aumentó 4%. El mayor incremento se dio en las secundarias Técnicas (5%), seguidas por las Telesecundarias y las Generales (4%). Los estudiantes en escuelas privadas aumentaron 2%. Las secundarias para trabajado-

res tuvieron una disminución en su matrícula de tercero de 17%.

Es interesante notar que no hay correspondencia entre las proporciones de aumento en la matrícula y en los planteles escolares. Esto se debe en buena medida a las grandes diferencias que existen en la cantidad de alumnos con que operan típicamente las escuelas de las diferentes modalidades.

Por otra parte, el número de escuelas y alumnos participantes en cada una de las dos muestras utilizadas para este estudio se pueden observar en la tabla II.

En la fase de planeación del estudio se decidió proporcionar resultados para las principales subpoblaciones en que se divide la educación secundaria en el Sistema Educativo Nacional: secundarias generales, técnicas, telesecundarias y

Tabla
I

Número de escuelas y matrícula en tercero de secundaria a nivel nacional y por tipo de servicio y sostenimiento en los ciclos 2004-2005 y 2007-2008

	Modalidad educativa	Escuelas				Alumnos			
		2005		2008		2005		2008	
		N	%	N	%	N	%	N	%
Tipo de servicio	Nacional	31208	100.0	33964	100.0	1782384	100.0	1856679	100.0
	General	9838	31.5	10641	32.5	907952	50.9	942484	50.9
	Técnica	4226	13.5	4432	13.5	495843	27.8	518235	28.0
	Telesecundaria	16800	53.8	17373	53.0	364933	20.5	380745	20.5
	Trabajadores	344	1.1	329	1.0	13656	0.8	11361	0.6
Sostenimiento	Federal, estatal o autónomo	27614	88.5	29866	87.9	1638911	92.0	1710105	92.1
	Privado	3594	11.5	4098	12.1	143473	8.0	146574	7.9

Tabla
II

Número de escuelas y alumnos en las muestras utilizadas para el estudio comparativo 2005-2008

Modalidad educativa	Escuelas				Alumnos			
	2005		2008		2005		2008	
	N	%	N	%	N	%	N	%
Nacional	2397	100.0	3304	100.0	52251	100.0	88198	100.0
General	566	23.6	881	26.7	16600	31.8	28896	32.8
Técnica	544	22.7	956	28.9	16093	30.8	31999	36.3
Telesecundaria	751	31.3	854	25.8	9720	18.6	13507	15.3
Privado	536	22.4	613	18.6	9838	18.8	13796	15.6

privadas. Se decidió excluir las modalidades Comunitaria y Para trabajadores, las cuales agregan a menos de 1% de la población objetivo. De igual forma se excluyeron aquellas secundarias que no se pudieron identificar con base en la información estadística de la SEP.

En el estudio de 2008 se decidió dar resultados según tres niveles de marginación para las dos modalidades que concentran más estudiantes (General y Técnica). Esto y las exigencias del diseño de comparación (que se comentan en el capítulo 1) causaron el incremento sustancial en el número de escuelas y alumnos en la muestra con respecto a la de 2005. Debido a que la desagregación por niveles de marginación no se consideró para el estudio de 2005, dicha división no puede considerarse para el trabajo comparativo de este reporte. Sobre los resultados de la aplicación 2008, pueden consultarse los capítulos 2 a 5 de *El aprendizaje en tercero de secundaria en México*. Informe sobre los resultados del Excale 09, aplicación 2008 (Sánchez y Andrade, 2009).

Adicionalmente, en cada una de las dos aplicaciones de la prueba se planeó dar resultados por entidad federativa y, al interior de las entidades, de las modalidades educativas en donde se concentra la mayor cantidad de alumnos. Esta segmentación de la muestra obliga a ser prudentes con los resultados por entidad: por un lado, se excluyen algunas modalidades en las tablas de resultados por entidad debido a que se refieren a cantidades muy pequeñas de alumnos; por otra parte, las comparaciones por entidad entre las dos aplicaciones se limitarán en este estudio a contrastar los casos en que hay diferencias significativas con la media nacional.

Para seleccionar a los alumnos en cada una de las dos muestras, en primer lugar se seleccionaron las escuelas participantes, y al interior de éstas se seleccionaron a los alumnos mediante procedimientos probabilísticos. Las escuelas se eligieron de forma sistemática a partir de listados ordenados de acuerdo con la cantidad de alumnos que atienden. Los estudiantes de cada escuela fueron seleccionados mediante un muestreo aleatorio simple. El carácter probabilístico permite hacer inferencias válidas para la población de alumnos del país y permite cuantificar el grado de precisión de los resultados. Haber utilizado el mismo procedimiento en ambas oca-

siones es uno de los elementos que apoyan la posibilidad de comparación.

Sobre el análisis de los resultados

En este informe, las escalas utilizadas para medir el logro se establecieron con el modelo de Rasch. Las puntuaciones obtenidas después fueron sometidas a una transformación lineal en la que:

- La media nacional es de 500 puntos.
- Una desviación estándar de la población nacional es igual a 100 puntos.
- Aunque los límites son abiertos, la mayoría de los valores posibles se ubican entre 200 y 800. Los valores cercanos a 200 indican menor grado de conocimientos y habilidades en la asignatura de que se esté hablando; los valores cercanos a 800 indican mayor grado de conocimientos y habilidades en la asignatura.

Cada asignatura tiene una escala independiente de las otras, lo cual significa que no pueden hacerse comparaciones directas entre una y otra; por ejemplo, asumiendo que la que tiene mayor puntuación promedio fue aprendida en mayor medida que la otra.

Las escalas para medir los resultados fueron fijadas en los valores señalados en los párrafos anteriores con la muestra nacional en la aplicación de 2005. De este modo, al obtener la media de grupos desagregados (por ejemplo, las modalidades), se pueden comparar las diferencias entre grupos. De la misma manera, se pueden comparar las diferencias entre años de aplicación.

Niveles de logro

Desde el reporte de resultados de la primera aplicación de los Excale se han utilizado los niveles de logro como una manera de facilitar la comprensión de los resultados en una visión global. Los niveles de logro utilizados para reportar los resultados de los Excale son cuatro y su descripción genérica se muestra en la tabla III.

Los niveles de logro son acumulativos conforme se va ascendiendo en ellos; es decir, los conocimientos y habilidades que tiene un estudiante que se encuentra en el nivel Por debajo del básico también los tiene un estudiante que se encuentra en el

nivel Básico, además de los que se describen en este nivel; de la misma manera, un estudiante en el nivel Medio tiene las habilidades y los conocimientos descritos en los dos niveles previos, además de los de su propio nivel. Finalmente, además de los conocimientos y las habilidades descritos en el nivel Avanzado, un estudiante que se encuentre en éste tiene también las habilidades y los conocimientos descritos en los tres niveles previos.

Estos niveles de logro tienen una expresión particular para cada una de las asignaturas y grados que se evalúan con los Excale, a la cual se llega a partir de un método específico, con la participación de especialistas de diferentes instituciones (Jornet y Backhoff, 2006). La expresión particular implica que se describen con detalle los diferentes conocimientos y habilidades que se tienen con respecto a una asignatura específica en cada nivel de logro. Los niveles de logro por asignatura se presentan al inicio de los capítulos correspondientes a los resultados del aprendizaje; a partir de ellos, los resultados pueden describirse en términos del porcentaje de alumnos que se encuentra en cada nivel.

En principio, los niveles de logro deben ser elementos fijos de la descripción de las escalas, de modo que puedan ser utilizados a lo largo de los años y ello permita la comparación en el tiempo de diferentes aplicaciones de los Excale a un mismo grado y asignatura. Tanto para la aplicación de 2005 como para la de 2008 se utilizaron niveles de logro delimitados por los mismos puntos de corte. Esto permite la comparación de ambas aplicaciones, aunque las definiciones de los niveles de logro utilizadas en 2008 están enriquecidas con respecto a las de 2005.



Antecedentes y estructura del informe

Este trabajo tiene tres antecedentes importantes en el INEE que se sugiere revisar para profundizar en la concepción que tiene el Instituto sobre la comparación del logro escolar a través del tiempo. En primer lugar, *La comparabilidad de los resultados de las pruebas nacionales de lectura y matemáticas 1998-2003* (Martínez Rizo, 2004), en el que se hace una breve pero completa revisión de los elementos que deben considerarse para hacer comparaciones de resultados educativos en el tiempo. En segundo lugar, el *Estudio comparativo de la educación básica en México: 2000-2005* (Backhoff *et al.*, 2005), en el cual se hace una reaplicación de la evaluación realizada por la Dirección General de Evaluación (DGE) de la SEP en el año 2000, ante la ausencia —en ese momento— de instrumentos diseñados para realizar comparaciones de rendimiento académico. Finalmente, el *Estudio comparativo del aprendizaje en*

Tabla
III

Niveles de logro genéricos

Nivel de logro	¿Qué indica?
Avanzado	Indica un dominio muy avanzado (intenso, inmejorable, óptimo o superior) de conocimientos, habilidades y destrezas escolares que reflejan el aprovechamiento máximo de lo previsto en el currículo.
Medio	Indica un dominio sustancial (adecuado, apropiado, correcto o considerable) de conocimientos, habilidades y destrezas escolares que pone de manifiesto un buen aprovechamiento de lo previsto en el currículo.
Básico	Indica un dominio imprescindible (suficiente, mínimo, esencial, fundamental o elemental) de conocimientos, habilidades y destrezas escolares necesarios para progresar satisfactoriamente en la asignatura.
Por debajo del básico	Indica carencias importantes en el dominio curricular de los conocimientos, habilidades y destrezas escolares que expresan una limitación para progresar satisfactoriamente en la asignatura.

sexto de primaria en México 2005-2007 (Backhoff *et al.*, 2008) utiliza por primera vez las pruebas Excale (concebidas desde su diseño para ofrecer comparaciones en el tiempo) de sexto de primaria para hacer un ejercicio de este tipo.

Antes de leer los capítulos de resultados, se sugiere revisar el capítulo 1 del reporte *El aprendizaje en tercero de secundaria en México*. Informe sobre los resultados del Excale 09, aplicación 2008 (Sánchez y Andrade, 2009) pues en él se hace una caracterización de las modalidades de educación secundaria a la cual se refiere este volumen.

Debido a que los reportes de posteriores aplicaciones incluirán tanto información a detalle de la aplicación vigente como datos comparativos con aplicaciones previas, este documento es el último que se presenta sólo con datos comparativos. Por ello, parece pertinente proponer en el capítulo 1 elementos para comprender las condiciones en que es posible hablar de comparabilidad entre pruebas para que sus resultados puedan ser analizados como mediciones sucesivas de un mismo constructo, así como los procedimientos específicos que se utilizaron para la comparación reportada en este informe y en otros comparativos de las pruebas Excale.

En el capítulo 2 se presenta brevemente la estructura del Excale 09 de Español, tanto en su aplicación 2005 como en su aplicación 2008; se presenta también la escala correspondiente y los niveles de logro con sus definiciones específicas para estas pruebas. Después se muestran comparaciones entre las dos aplicaciones, tanto para la población nacional como para algunos grupos en que las muestras pudieron desagregarse.



En el capítulo 3 se da una descripción similar a la del capítulo anterior, pero referida a la estructura de la prueba, la escala, los niveles de logro y los resultados de Matemáticas.

Finalmente, el capítulo 4 muestra una síntesis de resultados y las conclusiones.

Al final del documento se pueden encontrar las referencias bibliográficas y algunos anexos que detallan elementos del proceso de evaluación. Es importante mencionar que una gran cantidad de personas puso su talento y tiempo en el desarrollo, aplicación y análisis de las dos pruebas que se mencionan en este documento. A todos ellos les agradecemos (véase Backhoff *et al.*, 2006 y Sánchez y Andrade, 2009).

Procedimiento de equiparación

La comparabilidad de resultados 2005-2008

Uno de los principales objetivos del INEE es medir los cambios en el logro académico de los estudiantes mexicanos a través del tiempo. Para ello, es necesario que los resultados de las pruebas administradas en diferentes momentos tengan una escala en común para que las comparaciones que se realicen sean válidas.

En sí, el objetivo de la comparabilidad radica en trasladar las inferencias hechas a partir de los resultados de una prueba a los resultados de otra. De principio, podemos plantear al menos dos preguntas al respecto: ¿las inferencias correspondientes a una prueba se podrán aplicar con justicia a los resultados de otra? ¿Existen limitaciones?

La respuesta a estas preguntas dependerá de las características de las pruebas a comparar, así como de los métodos estadísticos implementados. Desde su construcción, cada prueba tiene sus propios propósitos y para lograrlos obtiene diferentes tipos de evidencia en magnitudes diferentes. Al final, a partir de toda la evidencia recolectada, se construyen inferencias.

En el campo de la evaluación educativa, la comparación entre resultados de las pruebas es un tema que desde principios del siglo pasado ha tomado especial interés. Los actores del proceso educativo quieren saber si los alumnos avanzan como lo marca el currículo, si los procedimientos de enseñanza han hecho posible la mejora de resultados, si los profesores han enseñado lo que se espera que enseñen, etc. Para dar respuesta a estas interrogantes es necesario medir el logro que los alumnos han alcanzado y llevar un regis-

tro de cómo este logro ha evolucionado. En otras palabras, es necesario comparar los resultados y construir tendencias.

El grado en que la comparabilidad de las pruebas puede tener éxito y la naturaleza de la maquinaria para llevarla a cabo dependen de la concordancia de los propósitos bajo los cuales éstas fueron construidas y de los aspectos de la competencia a mostrar para la que fueron diseñadas. Adicionalmente, es muy importante considerar la estandarización en la administración de las pruebas, entendiendo que ésta significa limitar las variaciones en la aplicación de las pruebas a las que los estudiantes están expuestos, en un intento de eliminar hipótesis sobre los resultados de los estudiantes que distraigan la atención de los propósitos planteados. Respecto a los propósitos de las pruebas, éstos pueden ser de alto o bajo impacto, referidos a norma o a criterio, y cuyos resultados competan a individuos o a grupos de individuos. Las concepciones que se tengan de competencia, ya sean conductuales, cognitivas o de otro tipo serán de vital importancia porque éstas definirán el concepto de pro eficiencia a utilizar en la interpretación de resultados y, por ende, repercutirán al momento de hacer comparaciones (Mislevy, 1992).

*Capítulo elaborado por Edgar Andrade Muñoz.

Como podrá observarse, la comparabilidad de pruebas no solamente abarca aspectos sobre técnicas estadísticas, sino también aspectos teóricos asociados al proceso de construcción de las mismas y los usos que van a tener. Por este motivo, cuando se comparan los resultados de pruebas es necesaria la identificación de diferencias y similitudes que orienten sobre el tipo de inferencias que se pueden o no trasladar, limitando los usos no adecuados o no justificados.

Para que el lector pueda seguir la racionalidad de los procesos de comparación de los Excale 09 de 2005 y 2008, a continuación se reseña parte de la experiencia de una de las instituciones de evaluación educativa de mayor trayectoria mundial. En segundo lugar, se realizan precisiones sobre el diseño de las pruebas de los Exámenes de la Calidad y el Logro Educativos y sus diferencias en las aplicaciones que se reportan en este documento. Por último, se explica la metodología utilizada para comparar sus resultados.

La enseñanza que dejó la anomalía de las pruebas de lectura de NAEP 1985-1986

En 1969 el programa NAEP (National Assessment of Educational Progress), en Estados Unidos, midió las tendencias de rendimiento académico de los alumnos en las áreas de Lectura, Escritura, Matemáticas y Ciencias, entre otras.

Para mejorar los procesos de medición y ofrecer al público reportes más atractivos, en 1984 y 1986 NAEP introdujo diversas modificaciones respecto a los nuevos conceptos curriculares y tecnologías de medición.

Para cuantificar el efecto de dichos cambios, se evaluó una muestra aleatoria de alumnos con las nuevas tecnologías y otra muestra aleatoria de alumnos, seleccionada con el mismo diseño muestral que la primera, con las mismas tecnologías y procedimientos de pruebas anteriores. Asumiendo que dichas muestras eran equivalentes en su nivel de competencia, los resultados de las evaluaciones anteriores fueron proyectados a la nueva escala de NAEP. Sin embargo, cuando se llevó a cabo la evaluación de Lectura en 1986, se realizaron pequeñas modificaciones en los cuadernillos de examen, así como en algunos procedimientos de administración de las pruebas sin la correspondiente muestra puente

(en inglés se emplea el término *linking*) para medir el efecto.

Los resultados mostraban que el logro de los alumnos era mucho menor respecto a los resultados de 1984, lo cual era muy improbable, dado que ocurrió en solamente dos años sin que los profesores en servicio lo advirtieran. El comité responsable de la publicación de resultados decidió que no debían ser usados para estimar las tendencias de la educación del país hasta que fueran respaldados por más evidencia.

Una vez que en 1988 se llevaron a cabo varios estudios al respecto, se mostró que el formato de los reactivos, su orden de aparición, las condiciones de tiempo, entre otros, cambiaban sustancialmente el comportamiento de dichos reactivos. Al volver a calcular las tendencias, se encontró que los resultados solamente habían bajado ligeramente respecto a 1984.

De este hecho surgieron varias lecciones para los procesos de evaluación en curso y los futuros. Albert Beaton, principal responsable de los estudios subsecuentes, señala que la mayor lección de todo el incidente es que cuando se midan los cambios no se debe cambiar el instrumento de medición, lo cual es imposible, como él mismo reconoce, pues NAEP ha modificado sus instrumentos desde que inició a medir las tendencias, debido a cambios sustanciales del currículo y a los avances en las tecnologías y metodologías de la evaluación. Se debe tener presente que el tamaño del efecto, debido a cambios en los instrumentos de medida, puede ser de tal magnitud que oculte los verdaderos cambios en el desempeño de logro de los alumnos, en especial cuando estos últimos son pequeños, como es común en la medición de tendencias. Beaton advierte que los cambios en la metodología de evaluación de tendencias son muy peligrosos y deben hacerse con mucha cautela (Beaton y Zwick, 1990).

Este caso, conocido como la anomalía de las pruebas de lectura de NAEP 1985-1986, muestra lo delicada que puede ser la implementación de mejoras en las pruebas destinadas a cuantificar tendencias en el logro académico de los alumnos a través de los años.

Además de los estudios realizados inmediatamente después de la anomalía, otros análisis han



mostrado que los cambios en el ordenamiento de los reactivos, el tipo de respuesta, el tipo de fuente de escritura y el formato de los cuadernillos pueden afectar las respuestas de los usuarios. A estos efectos de confusión los han llamado efectos de contexto y fundamentalmente se deben a que el supuesto de la independencia local de la Teoría de Respuesta al Ítem (TRI) no siempre se sostiene en los datos empíricos. Dicha independencia local se refiere a que la probabilidad de contestar cada reactivo en una prueba es independiente de la probabilidad de contestar los otros reactivos. Debido a que el contexto de una prueba siempre ejerce efectos de contexto sobre la estimación de los parámetros de los reactivos, éstos tienden a ser no robustos cuando hay cambios en las pruebas. Los parámetros obtenidos en cualquier configuración de cuadernillo de examen no son siempre apropiados para otras configuraciones concebibles (Qian, 2006).

Todo lo anterior hace, como bien lo señala Jiahe Qian del Educational Testing Service (ETS), que la medición de tendencias de logro sea una tarea difícil.

Una experiencia más reciente de NAEP. Las pruebas de Ciencias de 2005

En 2005, a las pruebas de Ciencias de NAEP se les hicieron dos cambios importantes respecto a las administradas en 2000: se modificó el formato de los cuadernillos y se substituyeron tres bloques de reactivos.

Para evitar efectos de confusión en las tendencias ocasionados por estos efectos de contexto, el nuevo diseño de 2005 agregó una muestra puente a la muestra principal del estudio (muestra operacional). A la muestra puente se le administraron los cuadernillos de examen en el mismo formato de la evaluación operacional de 2000. De esta manera, la evaluación puente y la previa fueron puestas en la misma escala mediante una calibración concurrente de reactivos y se usó una transformación lineal para vincular la escala resultante con la escala ya utilizada en los reportes. Además, debido a que la muestra operacional y la muestra puente (de 2005) compartieron la misma población, la escala de la muestra operacional pudo ser alineada con la escala de 2000 mediante la de la muestra puente ajustando el promedio y su desviación estándar de ambas distribuciones.

A juicio de Jiahe Qian, el diseño y análisis del NAEP de Ciencias de 2005 condujo a una revisión de la lección de Beaton: “para medir una tendencia con cambios de contexto, puentea los cambios pero no cambies el instrumento de medida (Qian, 2006)”.

El diseño matricial de las pruebas Excale 09: 2005 y 2008

A continuación se describen los diseños matriciales de los Excale 09 de 2005 y 2008. Se ha agregado información adicional para fundamentar su racionalidad, sus cambios y los análisis estadísticos implementados para obtener escalas de logro comparables entre ambos.

En 2004, cuando el INEE recién iniciaba sus labores, se solicitó el apoyo de expertos nacionales e internacionales sobre evaluación para que aconsejaran al Instituto sobre el diseño de las evaluaciones nacionales de logro académico que implementaría en un futuro cercano. Después de algunos meses de labor, el INEE ya contaba con un marco conceptual de los Excale (Backhoff y Díaz, 2005) sobre lo que medirían las pruebas y la forma de implementarlo, aspectos que se han expuesto de forma breve en la introducción de este libro.

Aplicación 2005

Una vez que se definió que las pruebas serían matriciales, se propuso que en 2005 se evaluarían las asignaturas de Español y Matemáticas. Por

su parte, los comités encargados del diseño de las pruebas definieron los contenidos a evaluar, así como sus reactivos asociados. Para Español, los contenidos se agruparon por una parte en aquéllos vinculados a la comprensión lectora y, por otra, aquéllos asociados a la gramática, ortografía, entre otros, que el programa de estudios engloba como Reflexión sobre la lengua. Dada esta distinción, los expertos invitados por el INEE para el diseño matricial recomendaron que dichos aspectos se midieran de forma separada, pues había sospechas de que estadísticamente se comportaran como dos dimensiones o variables latentes diferentes. El diseño propuesto permitiría en su

momento obtener datos para analizar ambos aspectos por separado o juntos. Así, para las pruebas de 2005, se evaluaron los siguientes aspectos:

- Comprensión de lectura
- Reflexión sobre la lengua
- Matemáticas

Los reactivos se agruparon en bloques que a su vez se integraron en cuadernillos de examen. Cada cuadernillo estuvo formado por dos bloques de cada uno de los aspectos evaluados. De esta manera cada alumno debía resolver seis bloques de reactivos: dos de Comprensión de lectura, dos

Tabla
IV

Bloques de reactivos para la aplicación 2005 del Exscale 09

Cuadernillo	Bloques					
	Lectura		Reflexión sobre la lengua		Matemáticas	
01	A	B	A	B	A	B
02	B	C	B	C	B	C
03	C	D	C	D	C	D
04	D	E	D	E	D	E
05	E	F	E	F	E	F
06	F	G	F	G	F	G
07	G	H	G	H	G	H
08	H	I	H	A	H	I
09	I	J	C	E	I	J
10	J	K	F	A	J	K
11	K	L	A	C	K	L
12	L	A	E	G	L	A
13	A	E	H	C	A	E
14	B	F	C	F	B	F
15	C	G	A	D	C	G
16	D	H	G	A	D	H
17	E	I	G	B	E	I
18	F	J	H	B	F	J
19	G	K	B	E	G	K
20	H	L	B	D	H	L
21	I	A	D	F	I	A
22	J	B	D	G	J	B
23	K	C	E	H	K	C
24	L	D	F	H	L	D

de Reflexión sobre la lengua y dos de Matemáticas. El tamaño de los bloques de reactivos se determinó con base en el tiempo necesario para que el alumno lo contestara con calma y para que los resultados no estuvieran influidos por el factor tiempo.

En total, para el Excale 09 de 2005, se construyeron 24 cuadernillos de examen como se muestra en la tabla IV.

Los cuadernillos de examen así contruidos se administraron a los alumnos de la muestra en espiral. Es como si todos los cuadernillos hubieran sido apilados siguiendo su orden: 1, 2, 3,..., 24, 1, 2, 3,... y así sucesivamente, para después ser repartidos a los alumnos en ese mismo orden. Este procedimiento permitió considerar que la muestra de alumnos que contestó cada tipo de cuadernillo fue, desde el punto de vista estadístico, equivalente.

Una vez que las pruebas fueron administradas se realizó el análisis de las mismas. Se comprobó que la asignatura de Español, que para efectos de la prueba se había dividido en dos aspectos, podía analizarse como una sola dimensión, lo cual permitía reestructurar a futuro las pruebas, dado que solamente medían Español y Matemáticas. Por su parte, los comités responsables del diseño de las pruebas revisaron los resultados obtenidos y decidieron mejorar las pruebas reformulando los reactivos con anomalías e incluyendo nuevos contenidos a evaluar para enriquecer las explicaciones sobre lo que los alumnos dominan del currículo.

Aplicación 2008

Para el Excale 09 a implantarse en 2008, el Plan de Evaluación del aprendizaje 2004-2016 del INEE

señalaba que se debían evaluar, además de Español y Matemáticas, las asignaturas de Biología y Formación cívica y ética.

Para la asignatura de Español, se planeó aumentar la cantidad de reactivos de la prueba en aproximadamente 36% y para Matemáticas 45%, como lo muestra la tabla V.

Si bien en un momento se pensó que el diseño matricial propuesto para 2005 serviría para los años siguientes, había varias dificultades técnicas que debían resolverse:

- Los cuadernillos de examen debían cambiarse necesariamente debido a la mejora de las pruebas de Español y Matemáticas que además de reestructurar reactivos agregaba otros nuevos.
- En los puntajes de Matemáticas de 2005 estaba incluido el efecto de cansancio de los alumnos —no cuantificado— por haber respondido antes reactivos de Español. Si bien la escala inicial de puntajes se centra en el valor 500 con una desviación estándar de 100 unidades, esto no resuelve el hecho de que la proporción de respuestas correctas de los reactivos de Matemáticas estuviera igualmente subvaluada en aplicaciones sucesivas.
- El tiempo de administración de las pruebas por alumno aumentaba al incrementar las asignaturas por evaluar, pues cada alumno debería contestar reactivos de las cuatro asignaturas.
- Se imposibilitaba la oportunidad de administrar cada una de las cuatro pruebas por separado y que los resultados fueran comparables, debido a que los puntajes inclui-

Tabla
V

Cambios en la cantidad de reactivos Excale 09 entre 2005 y 2008, por asignatura

Asignatura	Cantidad de reactivos	
	Prueba 2005	Prueba 2008
Español/Comprensión de lectura y reflexión sobre la lengua	113	149 (36 reactivos nuevos ~32%)
Matemáticas	128	186 (58 reactivos nuevos ~45%)
Formación Cívica y Ética	—	115
Biología	—	100

rían efectos de cansancio acumulados no cuantificados.

- Se tendrían que generar muchos cuadernillos para permitir que las combinaciones de bloques de reactivos fuera la mayor posible y que cada bloque estuviera al principio tantas veces como en segundo término (para controlar el efecto de cansancio).

Por estas razones se propuso cambiar el diseño matricial de las pruebas para 2008 por uno en que:

- Cada asignatura se evalúe de forma independiente. Es decir, cada alumno debe responder a reactivos de una sola asignatura.
- Cada cuadernillo de examen requiera por parte del alumno un tiempo de resolución equivalente al utilizado para responder los cuadernillos del diseño matricial de 2005, es decir, la carga para cada alumno deberá conservarse. De esta forma, los procedimientos de administración de las pruebas se conservan.
- La cantidad de cuadernillos por asignatura se reduzca sin que las interacciones entre bloques de reactivos demeriten.

En el documento *Technical Review of PISA* (Hambleton, Gonzalez, Plake y Ponocny, 2005), los autores propusieron cambiar el diseño matricial de las pruebas de PISA 2009 a uno similar a los implementados recientemente en NAEP. En dicho diseño, de bloques balanceado incompleto,

cada alumno resuelve reactivos de una sola asignatura y en cada cuadernillo se incluyen tres bloques de reactivos que se van rotando entre los cuadernillos de manera que cada uno de los bloques ocupe la primera posición, la segunda y la tercera en una secuencia en espiral. En la tabla VI se muestra un ejemplo para ocho bloques de reactivos, los cuales se acomodan en ocho cuadernillos de examen.

Lo anterior significa que por cada ocho alumnos que resuelvan las pruebas se tendrán tres respuestas por reactivo, una para cada posición de los bloques.

Después de valorar las virtudes de dicho arreglo, se decidió implementarlo en las pruebas Excale 09 de 2008, ya que respondía a los requisitos planteados, especialmente el correspondiente a su eficiencia, donde con menos cuadernillos de examen se obtiene mayor cantidad de interacciones entre bloques. La cantidad de interacciones es importante porque los modelos de medición a utilizar escalan los resultados de todos los cuadernillos como si se tratara de una sola prueba. Sin embargo, una desventaja es que se requiere aumentar el tamaño de las muestras para obtener errores asociados al muestreo de la misma magnitud que en otros arreglos.

En resumen, las pruebas de los Excale 09 de Español y Matemáticas administradas en 2005 y las administradas en 2008 tienen las siguientes similitudes y diferencias:

Tabla VI

Ejemplo de diseño matricial propuesto para Excale 09 aplicación 2008

Cuadernillo	Bloques de reactivos		
	1a. posición	2a. posición	3a. posición
1	A	B	E
2	B	C	F
3	C	D	G
4	D	E	H
5	E	F	A
6	F	G	B
7	G	H	C
8	H	A	D

Similitudes

- Comparten el mismo marco de referencia.
- Están diseñadas para obtener inferencias de la misma variable latente, es decir, que han sido construidas para proveer evidencia sobre la misma concepción de competencia.
- Los reactivos nuevos se construyeron con las mismas especificaciones técnicas.
- Están enfocadas en obtener inferencias de la misma población objetivo.
- Tienen los mismos procedimientos de administración, instrucciones, tiempos, etcétera.
- En ambos casos, los resultados de las pruebas son de bajo impacto en los alumnos, profesores y directivos, es decir, no hay incentivos asociados a prestigio, ingreso económico, etcétera.
- Comparten una cantidad considerable de reactivos.

Diferencias

- Las secciones de Comprensión de lectura y Reflexiones sobre la lengua se integraron.
- La prueba de 2008 es más extensa que la de 2005.
- Los reactivos cambiaron de posición.
- La estructura de la prueba cambió de manera significativa.
- El efecto acumulado de cansancio —por resolver reactivos de otra asignatura— se elimina en las pruebas de 2008.
- Las pruebas de 2008 requieren de una cantidad mayor de alumnos en la muestra.

Es importante notar que dadas las similitudes existentes, la tarea principal en la comparación de resultados se centra en encontrar la metodología que produzca los estadísticos de tendencia más robustos.

La comparación de Español y Matemáticas, 2005 y 2008

En 2005, cuando se administraron las pruebas de Español y Matemáticas del Excale 09 por primera vez, se obtuvieron los puntajes de los alumnos y se definió que la población nacional de alumnos tuviera un puntaje promedio de 500 unidades con una desviación estándar de 100 puntos para cada una de las asignaturas (Backhoff *et al.*, 2006). Es decir, se estableció la línea base a partir de la cual se compararían los resultados de los alumnos a través del tiempo. Si bien la escala es abierta por

los extremos, las puntuaciones oscilaron entre 200 a 800 unidades, aproximadamente.

Con base en las similitudes y diferencias de las pruebas de 2005 y 2008 expuestas anteriormente, así como la revisión de las experiencias acumuladas por NAEP, se llegó a la conclusión de que ambas pruebas debían compararse a través de una muestra puente utilizando dos procesos de vinculación (en inglés se emplea el término *linking*) de manera similar a la implementada por las pruebas de Ciencias de NAEP en 2005. La racionalidad de agregar una muestra puente fue la de asegurar la estabilidad de la medida de tendencias bajo un cambio en el diseño de cuadernillos y reducir los posibles errores por efectos de contexto debido a los cambios.

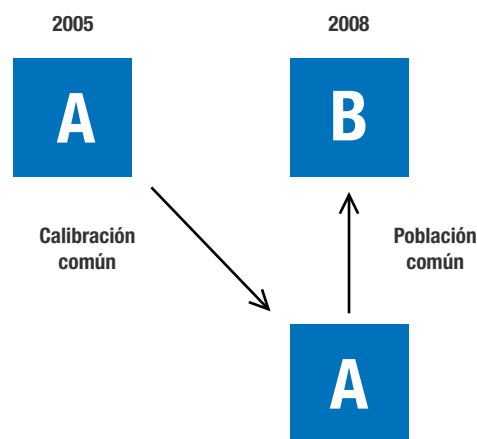
Se desechó la idea de comparar las pruebas simplemente mediante la equiparación por reactivos en común con base en un modelo de la Teoría de Respuesta al Ítem (TRI), debido a que dicho procedimiento asume que los reactivos miden exactamente el mismo constructo de dos o más poblaciones, a pesar de las diferencias en contexto y administración (Qian, 2006).

La figura 2 muestra la forma en que se realizó la comparación de resultados.

La muestra puente implementada en 2008 es equivalente pero de menor tamaño a la muestra operacional de ese mismo año desde el punto

Figura 2

Esquema para la aplicación y comparación de pruebas Excale 2005-2008



A: prueba diseñada para administrar en 2005
B: prueba diseñada para administrar en 2008

de vista estadístico, ya que es una submuestra aleatoria de la segunda. A esa muestra puente se le administraron las pruebas de 2005 bajo los mismos criterios de administración e idénticos cuadernillos de examen. Es decir, para medir el cambio de 2005 a 2008 se utilizó el mismo instrumento de medida.

Los procesos de vinculación fueron implementados en las dos etapas que a continuación se describen:

- Etapa 1. Mediante una calibración concurrente (TRI), la muestra puente y la evaluación de 2005 se pusieron en una escala en común y, utilizando una transformación lineal, la escala se ajustó para que coincidiera con la escala reportada ya anteriormente. Es decir, la muestra operacional de 2005 y la muestra puente de 2008 fueron calibradas juntas obteniendo puntajes para todos los alumnos evaluados en ambas muestras. Posteriormente, los puntajes de la muestra de 2005 fueron transformados mediante una ecuación lineal de tal forma que su promedio fuera de 500 con una desviación estándar de 100 unidades como se hizo en 2005 cuando se definió la escala de reporte. Utilizando la misma transformación lineal, los puntajes de la muestra puente (de 2008) fueron transformados a la escala de 2005 pudiendo estimar el avance de los alumnos en esos tres años.
- Etapa 2. La escala de la muestra puente, ya con la misma escala de los reportes anteriores, se trasladó a la muestra operacional de 2008 a partir del hecho de que comparten la misma población. Es decir, una vez que la muestra operacional de 2008 se calibró independientemente, su promedio y su desviación estándar se ajustaron mediante una transformación lineal para que coincidiera con los de la muestra puente que se obtuvo al finalizar la etapa 1.

Antes de implementar la calibración concurrente entre la muestra puente y la muestra operacional de 2005, se verificó que el comportamiento de los reactivos fuera similar. Aunque se administró la misma prueba de forma idéntica, siempre es posible que algunos reactivos tengan un comportamiento atípico debido a causas ajenas a la implementación del currículo, como pueden ser campañas televisivas, temas que se ponen de moda, etc. Antes de



llevar a cabo la calibración conjunta, cada una de las muestras se calibró de forma separada y las dificultades de los reactivos se graficaron para detectar estos casos atípicos. Solamente cuatro de los reactivos (dos de Español y dos de Matemáticas) se ubicaron fuera de las bandas de confianza de 95% de la recta de regresión, por lo que fueron excluidos de la calibración conjunta.

Respecto a la etapa 2, en la cual se realizó la vinculación de la escala considerando que la muestra puente y la muestra operacional de 2008 tenían a la población en común, se compararon los percentiles de ambas distribuciones una vez que se realizó el traslado de la escala. Es de esperar que, teniendo a la población en común, ambas muestras tengan la misma distribución y por ende los mismos percentiles. En la tabla VII se muestran los percentiles calculados para la muestra puente y la muestra operacional de 2008. Como se observa en la tabla, en ningún caso la diferencia entre los percentiles es estadísticamente significativa, es decir, no se cuenta con evidencia para señalar que las distribuciones de alumnos sean distintas.

En suma, la experiencia internacional ha mostrado que la comparación de resultados con la finalidad de obtener tendencias a través del tiempo es una tarea muy delicada y difícil. Por su parte, el INEE ha incorporado las experiencias sobre medición de tendencias de una de las instituciones de evaluación educativa con mayor trayectoria en el mundo con la finalidad de obtener resultados robustos y que sean estables.

Cuantiles de la prueba puente y muestra operacional Excale 09 aplicación 2008

Asignatura	Percentil	Muestra puente		Muestra operacional 2008		Diferencia		
		Valor	(EE)	Valor	(EE)	Valor	(EE)	t
Español	10	369	4.7	374	2.2	-4.3	7.1	-0.61
	25	430	4.2	426	2.1	4.2	4.6	0.93
	50	497	3.4	491	2.6	6.0	6.0	0.99
	75	564	3.6	566	2.9	-2.1	5.4	-0.39
	90	626	4.9	633	3.5	-6.9	6.0	-1.15
Matemáticas	10	381	4.5	383	1.8	-2.8	4.8	-0.58
	25	436	4.0	433	1.8	3.3	5.3	0.63
	50	499	3.2	495	2.0	4.4	4.5	0.98
	75	567	4.5	566	2.9	1.2	7.0	0.17
	90	638	5.6	640	4.1	-1.9	8.0	-0.24

El error estándar de los cuantiles fue obtenido por el método de remuestreo Jackknife (JKn). EE corresponde a los errores estándar asociados a la prueba t (de student) en todos casos tiene más de 100 grados de libertad.

Las pruebas Excale 09 de Español y Matemáticas administradas en 2008 tuvieron dos modificaciones respecto a la administradas en 2005: 1) incorporaron nuevos contenidos curriculares a evaluar y 2) cambiaron el diseño matricial de las pruebas. En cuanto a su construcción, ambas pruebas comparten el mismo marco de referencia, fueron diseñadas para mostrar evidencia sobre los mis-

mos conjuntos de aprendizajes, están enfocadas a obtener inferencias sobre la misma población objetivo, sus resultados tienen el mismo bajo impacto y comparten los mismos procedimientos de administración. Estas similitudes justifican que las inferencias realizadas en 2005 puedan trasladarse a los resultados de 2008 una vez que los resultados son puestos en una escala de medida común.

Resultados y comparación en Español

Definición de la escala de Español¹

El INEE evaluó en 2005 por primera vez la asignatura de Español. Los Excale 09 retomaron los contenidos vinculados con los propósitos específicos mencionados por la SEP factibles de ser evaluados que, en el caso de la asignatura de Español, fueron que los alumnos:

- Desarrollen conocimientos y estrategias para comprender distintos tipos de texto.
- Se formen como lectores que valoren críticamente lo que leen.
- Desarrollen conocimientos y habilidades para buscar, seleccionar, procesar y emplear información como aprendizaje autónomo.
- Logren comprender el funcionamiento y las características básicas de nuestro sistema de escritura de manera eficaz.
- Desarrollen estrategias para comprender y ampliar su lenguaje, así como para escribir.
- Adquieran nociones de gramática para que puedan reflexionar y hablar sobre la forma y uso del lenguaje escrito como un recurso para mejorar la comunicación.

En la elaboración de los reactivos utilizados en el Excale se hizo hincapié en respetar el enfoque comunicativo y funcional.

Debido a que el plan de evaluación del Instituto se basa en un programa cuatrianual (un mismo grado escolar se evalúa cada cuatro años),

en 2008 se volvió a evaluar este grado escolar. Con este motivo se mejoró la calidad técnica de la prueba, para lo cual se realizó una revisión del Excale empleado anteriormente.

Para la revisión se decidió alinear el Excale a los programas de estudio de 1993 por última vez, debido a que los estudiantes a evaluar fueron la última generación que se formó con los planes y programas de ese año. Para próximos ejercicios de comparación entre años se tendrá que hacer uso de estrategias distintas para lograr que las pruebas sean comparables entre sí.

En el rediseño de la prueba de Español se respetaron los siguientes criterios:

- Alinear la prueba al Plan y Programas de estudio 1993.
- Utilizar la mayor parte de los reactivos aplicados en 2005, eliminando o modificando sólo un pequeño porcentaje —aquéllos que mostraron alguna dificultad en su validez de contenido—.
- Adicionar reactivos que evaluaran contenidos no examinados en 2005.

Con el fin de comunicar los resultados de la manera más eficiente y lógica posible, los contenidos

¹ Este apartado es una síntesis de la sección correspondiente en Sánchez y Andrade (2009), cuya autoría corresponde a Margarita María Peón Zapata. Para profundizar en las características de la prueba y de la escala de Español se sugiere revisar el documento original.

Líneas de evaluación	Habilidades y conocimientos	Reactivos 2005	Reactivos rediseño
Comprensión lectora 63%	Extracción de información	3	-
	Desarrollo de una comprensión global	19	6
	Desarrollo de una interpretación	23	7
	Análisis del contenido y la estructura	18	8
	Evaluación crítica del texto	3	-
Reflexión sobre la lengua 37%	Reflexión semántica	13	3
	Reflexión sintáctica y morfosintáctica	14	8
	Reflexión sobre las convencionalidades de la lengua	13	-
Subtotal		106	32
Total		138	

se clasificaron en siete categorías; esto no significa que un lector aplique estas habilidades de manera seriada —primero una, después otra—. La clasificación resulta oportuna sólo para comprender los resultados más fácilmente; en este sentido, la tabla VIII muestra esta clasificación, así como la cantidad de reactivos utilizados para evaluar cada categoría.

Para la prueba revisada se tomaron 103 reactivos de los 108 aplicados en 2005. Dos de los restantes se eliminaron y tres se reformularon para mejorar su validez. Por otra parte, se introdujeron 32 reactivos nuevos para evaluar contenidos no tomados en cuenta en la prueba.

Como es habitual en las pruebas Excale, se desarrollaron niveles de logro específicos que se presentan en la tabla IX. La población evaluada se distribuye en los cuatro niveles, y para cada uno de éstos se enuncian las habilidades y conocimientos curriculares que han mostrado tener los alumnos que se encuentran en él. Debe tomarse en cuenta que los niveles son acumulativos, de modo que los estudiantes que se encuentran en el nivel Avanzado han mostrado que tienen las habilidades y los conocimientos curriculares de todos los niveles previos, los del nivel Medio tienen las habilidades y los conocimientos de los niveles Básico y Por debajo del básico, etcétera.

Con la finalidad de dar significado práctico a la escala de resultados de aprendizaje en Español, se presenta la figura 3, la cual muestra un mapa parcial de reactivos²; cabe hacer notar que en esta figura sólo se muestran algunos reactivos, por lo que los contenidos más fáciles y más difíciles no necesariamente son los que aquí aparecen. El mapa divide, con líneas punteadas tenues, los tres puntos de corte, a partir de los cuales se agrupan los reactivos en los cuatro niveles de logro. La columna de la izquierda nos muestra la escala, que va de 200 a 800 puntos. En la columna central se describen algunos reactivos de la prueba, cada uno de ellos asociado a su dificultad. Por último, en la columna de la derecha aparece el porcentaje de alumnos que acertaron el reactivo.

Resultados comparativos en Español

Los resultados descritos de manera extensa para cada una de las dos aplicaciones que se utilizan en este estudio se presentan en reportes detallados realizados previamente por el INEE. Para el caso de los datos de la aplicación 2005, los resultados se encuentran en *El aprendizaje del Español y las Matemáticas en la educación básica en México* (Backhoff et al., 2006). Los datos de la

2 El mapa parcial se refiere a la escala de la aplicación más reciente.

Tabla
IX

Definición de los niveles de logro en Español

Nivel de logro

Avanzado

En *Comprensión lectora* los alumnos ubicados en este nivel interpretan la información que proviene de diferentes partes del texto en ensayos y textos científicos, así como el contenido y organización de las ideas, argumentos, contra-argumentos y elementos persuasivos. Reconocen las oraciones temáticas de artículos de opinión y de divulgación, así como la trama de cuentos clásicos. Pueden identificar el punto de vista del autor en algunos sonetos de Sor Juana Inés de la Cruz y en textos satíricos. En estos últimos pueden comprender el sentido real del mensaje e identificar su tono irónico. En poemas interpretan imagen y metáfora e identifican ritmo y métrica. Relacionan la información del texto con la que se ofrece en las imágenes, tablas y gráficas. Evalúan críticamente prejuicios en artículos de opinión y reportajes.

En *Reflexión sobre la lengua* los alumnos reconocen el significado y el uso correcto de palabras homófonas con y/lI y s/c/z. Relacionan una sustitución léxica con el nombre al que reemplazan. Reconocen las palabras y frases que establecen relaciones de coherencia y cohesión en un texto (marcadores discursivos). Son capaces de identificar oraciones yuxtapuestas y coordinadas, así como la función de la oración subordinada en contexto. Reconocen el uso correcto de verbos en gerundio, así como de los verbos conjugados. Identifican la función y uso del modo imperativo. Reconocen el sujeto morfológico entre tres o más posibles sujetos explícitos en el texto.

Medio

En *Comprensión lectora* los alumnos ubicados en este nivel reconocen el propósito de reportajes, así como el tema central de cuentos de ficción. Pueden sustituir el mensaje de estos cuentos por un refrán o proverbio, así como precisar el contexto sociocultural, el espacio y ambiente donde ocurren los hechos, además de valorar la congruencia entre personaje/caracterización, acción/lugar. Interpretan adecuadamente el sentido irónico expresado en una crónica. Identifican la secuencia argumentativa de un texto científico, las relaciones causa/efecto y problema/solución en textos informativos, así como conclusiones expresadas en paráfrasis; identifican las circunstancias que rodean a un hecho (noticia). En cuanto al análisis de contenido y la estructura, pueden evaluar si las ideas están organizadas de manera adecuada según el tipo de texto. Interpretan correctamente la información de una gráfica; reconocen las partes de una carta comercial; identifican el orden lógico de las instrucciones, la organización y secuencia de las noticias, así como la estructura de una monografía; detectan el clímax de una leyenda, el desarrollo y la conclusión de un ensayo e identifican las características formales de un poema tradicional.

En *Reflexión sobre la lengua* los alumnos ubicados en este nivel infieren el significado de expresiones metafóricas. Reconocen el uso adecuado de palabras homófonas con b/v. Detectan errores en el uso del participio y de concordancia verbal en un texto. Identifican los nexos que introducen una secuencia de ideas. Reconocen oraciones subordinadas. Interpretan el sentido de una oración por su puntuación e identifican su uso correcto en oraciones yuxtapuestas. Reconocen el uso de la coma para organizar elementos de la oración, así como la función de las comillas, guiones largos, puntos suspensivos, signos de interrogación y de admiración. Detectan errores ortográficos en palabras con ger/gir, gue/gui, ble, bilidad, bir, aba, vo, va.

Básico

En *Comprensión lectora* los alumnos pertenecientes a este nivel reconocen el propósito del ensayo, la entrevista y de la carta comercial, así como el tema central de crónicas, monografías, comentarios, obras de teatro, cuentos tradicionales y fábulas. En estas últimas pueden sustituir el tema por un refrán. Pueden identificar las intenciones implícitas y características de los personajes de un cuento o fábula. Identifican la función y la estructura de los instructivos y recetas, así como las características de las crónicas. Reconocen la opinión del autor y las conclusiones que se pueden derivar de un reportaje, ensayo o artículo de opinión. Pueden completar un cuadro sinóptico con la información obtenida de un texto breve, así como identificar la idea principal de un párrafo, la secuencia temporal y las relaciones causa/efecto en textos narrativos. Detectan la organización de las ideas dentro de un ensayo y una entrevista.

En *Reflexión sobre la lengua* los alumnos de este nivel pueden interpretar el significado de palabras, incluyendo las indígenas de uso frecuente, con ayuda del contexto, así como inferir el sentido denotativo de una expresión. Pueden sustituir una palabra que aparece en un texto por un sinónimo. Reconocen una función del infinitivo e interpretan correctamente los accidentes del verbo. Reconocen el uso adecuado de las conjunciones. Identifican las palabras que deben llevar acento diacrítico y reconocen el uso correcto de la acentuación en palabras agudas, graves y esdrújulas.

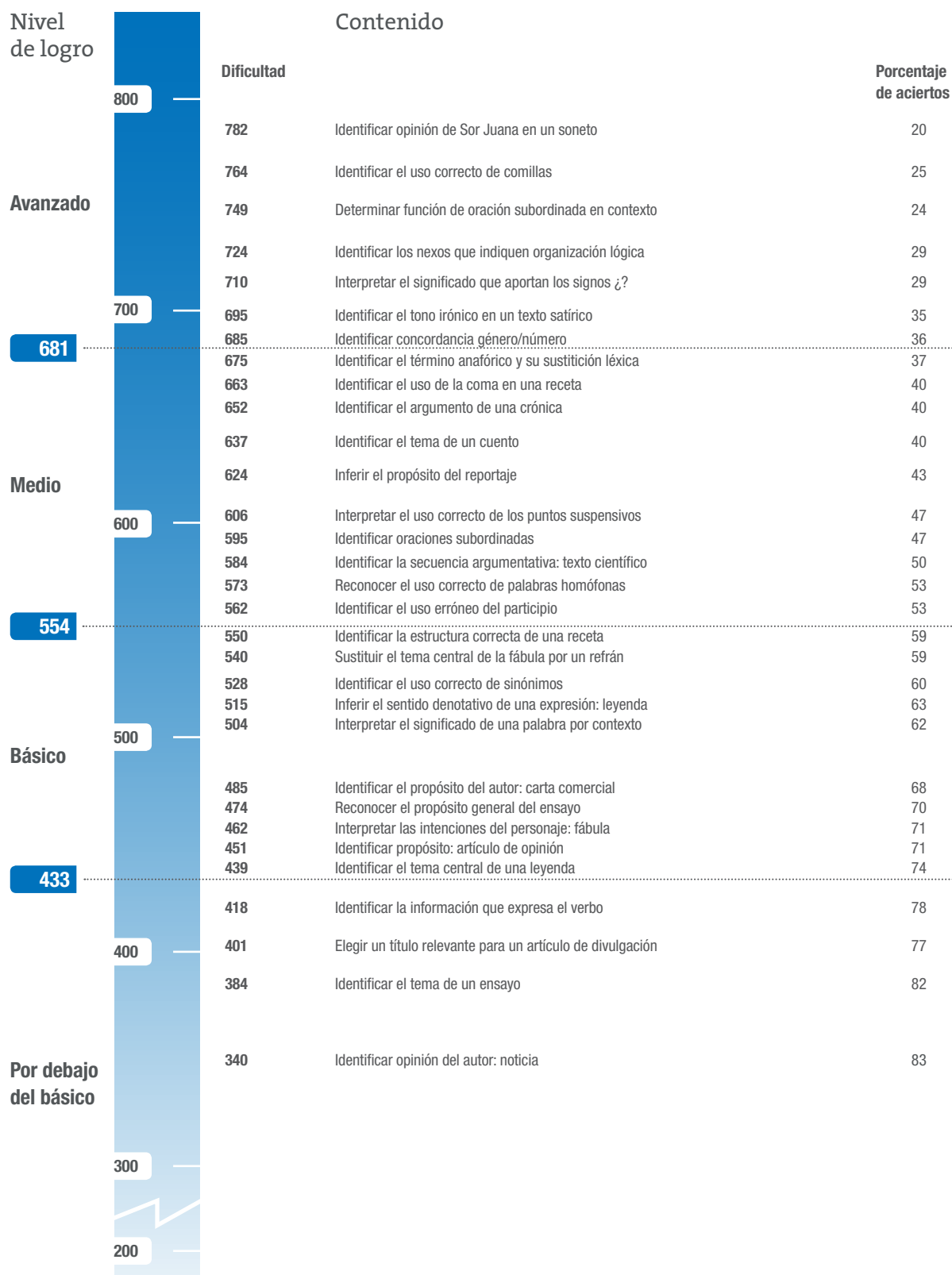
Por debajo del básico

En *Comprensión lectora* los alumnos ubicados en este nivel pueden identificar el propósito y el tema de textos con un contenido familiar, así como seleccionar un título relevante que los encabece. Pueden identificar las características gráficas y/o contenido de leyendas e instructivos.

En *Reflexión sobre la Lengua* los alumnos de este nivel detectan los objetos directo e indirecto; identifican los verbos: *ser, estar, parecer, lucir, quedar y llegar* que pueden completar un texto. Reconocen el uso adecuado de preposiciones e identifican el uso correcto de “h” (en comienzo de palabras con sílabas -hie, -hue, -hum).

Figura
3

Mapa parcial de reactivos: Español³



³ Si el lector desea conocer la información de la totalidad de los reactivos que conformaron esta prueba, puede consultar el Anexo 1 de Sánchez y Andrade (2009).

aplicación 2008 fueron trabajados a profundidad en *El aprendizaje en tercero de secundaria en México. Informe sobre los resultados del Excale 09, aplicación 2008* (Sánchez y Andrade, 2009).

Comparación por modalidad

Una de las preocupaciones centrales en los análisis de resultados de los Excale ha sido el monitoreo de las diferencias en el logro educativo entre modalidades educativas. Esto se debe a la importancia que tiene el conocer si la escuela en sus diferentes servicios educativos está contribuyendo a disminuir la desigualdad entre grupos o a aumentarla. Las distintas modalidades educativas centran su atención en poblaciones distintas, en términos del contexto en que se ubican sus escuelas, sus condiciones económicas y sus antecedentes escolares (*cfr.* Santos, 2009 en: Sánchez y Andrade, 2009) y, sin embargo, tienen propósitos curriculares comunes. ¿En qué medida alcanzan las diferentes modalidades estos propósitos? Esta es la pregunta central para este apartado.

En la tabla x se pueden observar las puntuaciones medias obtenidas en 2005 y 2008 tanto para la población nacional como para las diferentes modalidades educativas consideradas. Las últimas columnas muestran la diferencia entre las puntuaciones de un año y otro, y su correspondiente error estándar.

Como puede observarse, sólo hay diferencias significativas entre años en la modalidad educativa de Telesecundaria. En los demás casos, no se cuenta con datos que evidencien cambios en el desempeño promedio de los estudiantes en Español.



La figura 4 muestra los mismos datos colocados en una gráfica. Al igual que en algunas gráficas subsecuentes, la media se señala con la figura en blanco dentro de cada rectángulo. Los límites inferior y superior de cada rectángulo indican el intervalo de confianza. Además se señalan los niveles de logro en el margen derecho y sus respectivos puntos de corte con líneas punteadas tenues que atraviesan horizontalmente la gráfica.

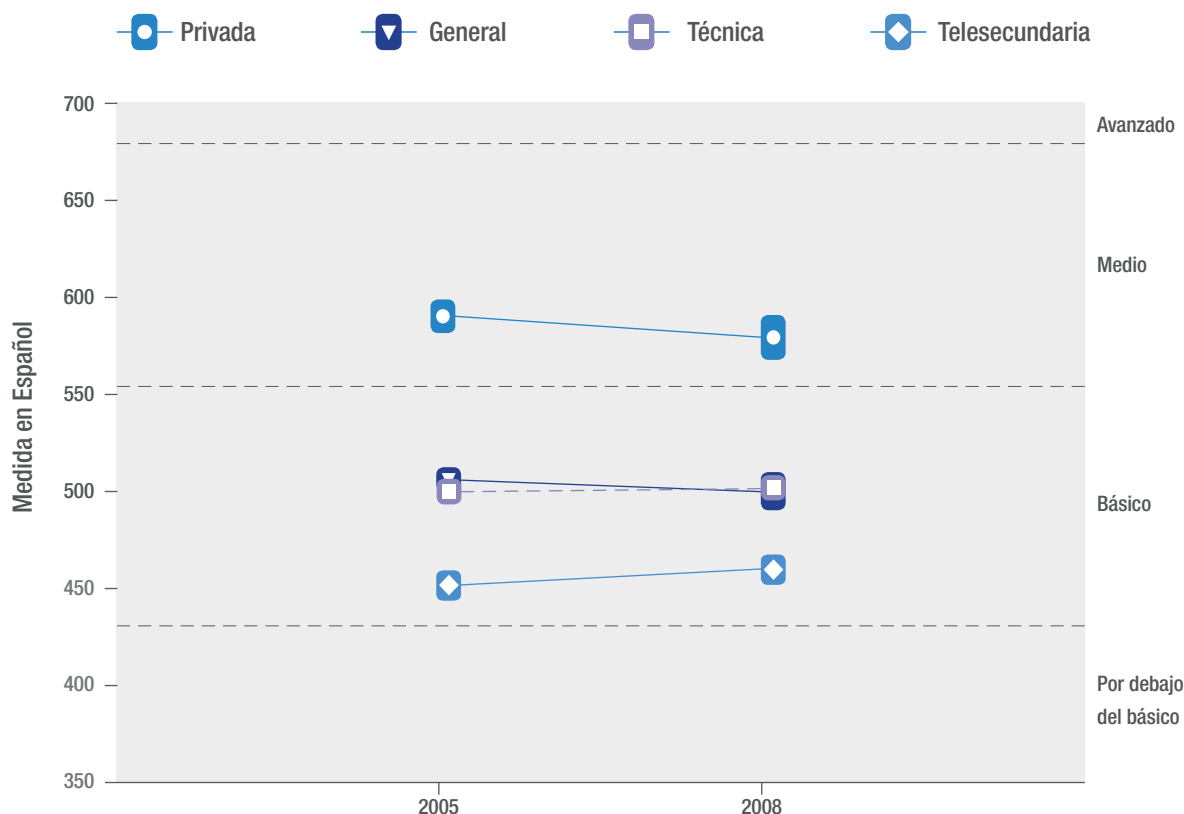
Ya se señalaba en los reportes específicos para describir los resultados de 2005 y de 2008 que, entre las modalidades educativas estudiadas, la Telesecundaria obtuvo resultados más bajos en ambas aplicaciones. Sin embargo, al observar los dos puntos simultáneamente, se puede ver que en los demás casos hay una pequeña tendencia a disminuir en la puntuación promedio, o a mantenerse, mientras que la Telesecundaria cambió de manera significativa de una puntuación promedio que estaba en el nivel de logro Por debajo del básico a otra que está en el nivel Básico.

Tabla
X

Puntuaciones medias por modalidad educativa en Español

Modalidad educativa	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)
Nacional	500	(1.3)	498	(2.0)	-2	(2.4)
General	505	(2.7)	500	(4.2)	-5	(5.0)
Técnica	502	(2.0)	502	(2.1)	0	(2.8)
Telesecundaria	452	(2.4)	460	(2.5)	8	(3.2)
Privada	589	(2.8)	578	(4.9)	-11	(5.6)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.



La figura 5 muestra de qué manera se distribuyen los estudiantes, tanto a nivel nacional como por modalidades educativas, en los cuatro niveles de logro descritos anteriormente, tanto en 2005 como en 2008. Puede observarse también en la columna “Matrícula” el porcentaje de alumnos que pertenecen a cada modalidad educativa, y en la de “Puntaje promedio” los valores medios que ya se mencionaron más arriba.

Resulta interesante observar las variaciones en los niveles de logro entre la aplicación de 2005 y la de 2008. En principio, es posible que la puntuación promedio no cambie, pero sí cambie la proporción de estudiantes en los diferentes niveles de logro. A pesar de que la puntuación media a nivel nacional se conserva prácticamente igual, en 2008 se tiene un 3.2%⁴ más de estudiantes

en el nivel Por debajo del básico y 0.8% más de estudiantes en el nivel Avanzado, así como 2.4 y 1.6% menos en el nivel Básico y en el nivel Medio, respectivamente.

Es importante tomar en cuenta que las estimaciones de porcentaje de alumnos por nivel de logro no son exactas, por lo que las consideraciones de “diferencias” entre años, sobre todo en los casos en que éstas son pequeñas, pueden deberse al error de medida (que siempre está presente) y no a cambios en el logro de los estudiantes.

En el caso de telesecundarias, en 2008 hay una disminución de 1% en el nivel Por debajo del básico y un aumento similar en el nivel Medio. Es interesante resaltar cómo un cambio de 1% entre estos dos niveles, que puede parecer muy modesto, participa en la ya mencionada variación significativa de 2005 a 2008 de los puntajes promedio.

En las secundarias técnicas hay aumentos de 2.6 y 1.5% en los niveles de logro Por debajo del básico y Avanzado respectivamente, en 2008 con respecto a 2005, acompañado por disminucio-

⁴ Los valores presentados en las tablas y gráficas de los capítulos están redondeados a números enteros para facilitar su lectura. Esto lleva a aritméticas inexactas. Por ejemplo, al sumar los porcentajes de algunas modalidades por año el resultado es ligeramente mayor al 100%. Los valores con decimales a partir de los cuales se hicieron los cálculos de diferencias entre porcentajes se encuentran en la tabla A del Anexo 1.

nes de 3 y 1.3% en los niveles Básico y Medio, respectivamente.

En las secundarias generales se observan aumentos de 5 y 1.1% en los niveles de logro Por debajo del básico y Avanzado, respectivamente, acompañados por disminuciones de 3.9 y 2.2% en los niveles Básico y Medio, respectivamente.

En la modalidad Privada se pueden ver aumentos en los porcentajes de alumnos en los niveles Por debajo del básico y Básico (3.8 y 0.5%, respectivamente), así como una disminución de 3.4 y 1% respectivamente en los niveles Medio y Avanzado.

Brechas entre modalidades

Considerar las brechas entre modalidades al hacer ejercicios de evaluación del logro como el presente permite monitorear si las modalidades que tienen menores puntajes promedio se van acercando en puntuación a las demás, lo cual apunta a una mayor equidad en el logro de los estudiantes.

Idealmente, los estudiantes con distintas condiciones de vida y en diferentes modalidades educativas deberían tener niveles de logro similares, y éstos deberían ser altos. Aún estamos lejos de ello.

Tomando nuevamente en cuenta los datos de la tabla X, se puede observar que la diferencia entre la modalidad con puntaje promedio más bajo (telesecundarias) y la modalidad con puntaje promedio más alto (privada) disminuye de 2005 (137 puntos) a 2008 (118 puntos).

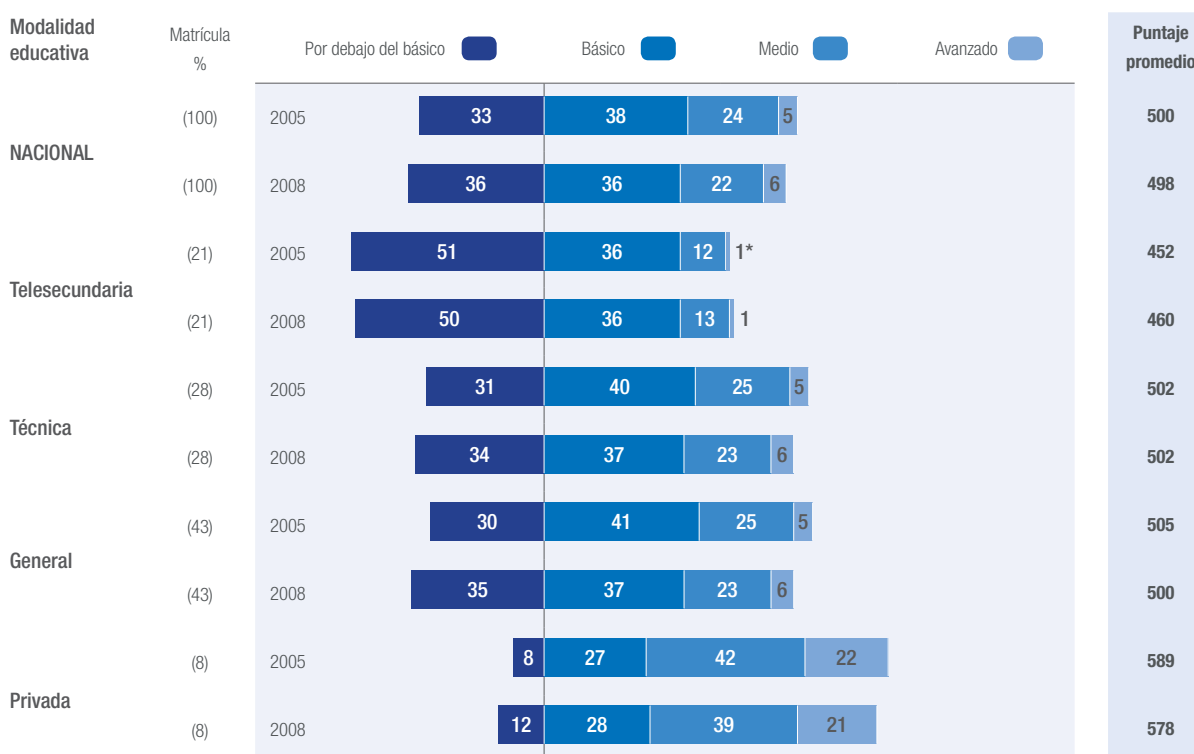
De la misma manera, la diferencia entre la modalidad con promedio más bajo (Telesecundaria) y la media nacional ha disminuido de 48 puntos en 2005 a 38 puntos en 2008.

Dispersión dentro de las modalidades

Además de monitorear las diferencias entre modalidades, es importante considerar las distancias entre los individuos con resultados altos y bajos al interior de los grupos. Mientras mayores sean éstas, mayor es la dispersión en los

Figura
5

Porcentaje de estudiantes por nivel de logro educativo y modalidad educativa en Español



* Estimación cuyo coeficiente de variación excede a 20%.

Porcentaje de estudiantes

grupos, lo que indica diferencias mayores en el logro escolar.

La tabla XI muestra dos medidas de la dispersión para la población nacional y las diferentes modalidades. La desviación estándar (DE) nos indica el grado de dispersión de los datos con respecto a la media, utilizando como unidad de medida la puntuación del Excale. El coeficiente de variación (CV) indica el porcentaje del tamaño de la desviación estándar con respecto a la media.

Como puede observarse, las diferencias en la dispersión de puntajes en Español no varían mucho entre las aplicaciones 2005 y 2008. Sólo el coeficiente de variación, en los casos de modalidad Técnica y Telesecundaria tienen cambios significativos; en el primer caso, muestra un ligero aumento en la dispersión de los puntajes dentro del grupo, mientras que en Telesecundaria hay una disminución de dos puntos porcentuales con respecto a la media en cuanto a la dispersión de los puntajes dentro del grupo, lo que indica que en la modalidad Técnica aumentaron ligeramente las diferencias en el logro escolar, mientras que en Telesecundaria disminuyeron ligeramente.

Comparación por sexo

La desagregación de los resultados por sexo tiene en principio una función de equidad. Monitorear si tanto alumnas como alumnos que han ingresado a la escuela alcanzan niveles de logro similares es de interés. Además de ello, algunas asignaturas han mostrado en repetidas ocasiones que mujeres y hombres tienen ligeras diferencias a favor de unos u otros en su desempeño. Tal es el caso



del Español, en que muchos estudios, incluyendo los Excale que se han realizado a la fecha, las mujeres muestran resultados promedio significativamente mejores que los hombres, aunque estas diferencias usualmente son pequeñas. Entre dos aplicaciones sucesivas de una misma prueba lo esperado sería seguir encontrando distancias similares en cada nueva aplicación.

La figura 6 muestra el comportamiento de las medias de mujeres y hombres en las aplicaciones 2005 y 2008, con sus respectivos intervalos de confianza.

En los reportes correspondientes —ya se ha mencionado que tanto en 2005 como en 2008— la diferencia en puntuaciones medias entre mujeres y hombres es significativa.

Al observar las dos aplicaciones en su conjunto, se puede advertir que las mujeres tienen una media tres puntos mayor en la segunda aplicación. Esta diferencia no es significativa. En cam-

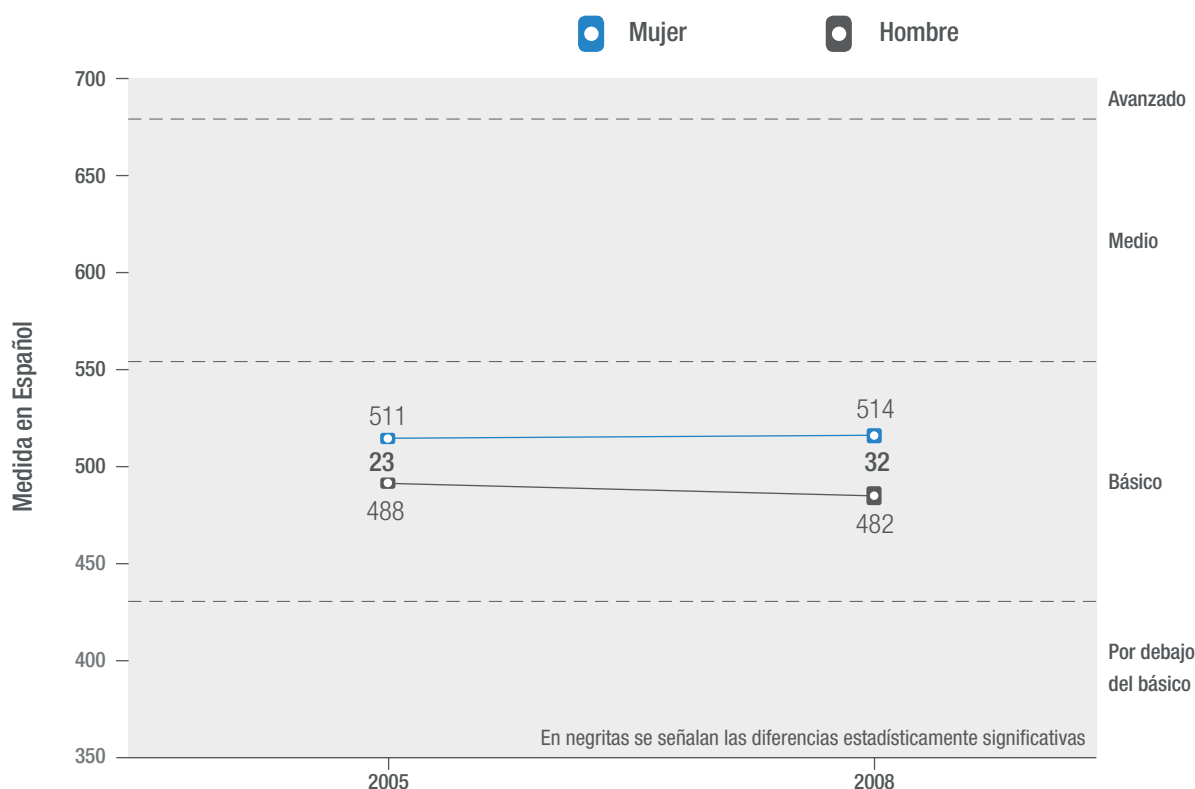
Tabla XI

Dispersión de puntajes en Español por modalidad

Modalidad educativa	2005				2008			
	D.E.	(EE)	C.V.	(EE)	D.E.	(EE)	C.V.	(EE)
Nacional	100	(0.7)	20	(0.1)	100	(1.0)	20	(0.2)
General	94	(1.1)	19	(0.2)	97	(1.6)	19	(0.3)
Técnica	96	(1.1)	19	(0.2)	99	(1.4)	20	(0.3)
Telesecundaria	93	(1.5)	21	(0.3)	87	(1.6)	19	(0.4)
Privada	95	(1.2)	16	(0.2)	101	(2.3)	17	(0.5)

Figura 6

Comparación de las puntuaciones medias en Español por sexo



bio, los hombres muestran una disminución de siete puntos⁵ en la segunda aplicación que sí es estadísticamente significativa. La distancia entre mujeres y hombres, por consiguiente, aumentó, y la diferencia entre las brechas de 2005 y 2008 también es significativa⁶.

La figura 7 muestra los resultados en Español para las dos aplicaciones, desagregados por sexo y expresados en términos de porcentajes de estudiantes que se encuentran en cada nivel de logro.

De 2005 a 2008 se muestra un aumento de cerca de 6 % en la proporción de hombres en el nivel Por debajo del básico⁷ correspondiente a una disminución en los niveles Básico y Medio, de poco más

de 3 y 2%, respectivamente. Con respecto a las mujeres, las proporciones en cada nivel de logro son prácticamente las mismas en los dos años.

Comparación por sexo y modalidad

Al desglosar la información de resultados por sexo en modalidades educativas, encontramos los resultados expresados en la tabla XII. Los datos en negritas corresponden a las diferencias significativas. Como puede observarse, además de la diferencia significativa ya mencionada en la media de los hombres a nivel nacional, hay otros dos grupos que tienen diferencias significativas: los hombres en la modalidad Privada, que disminuyeron 25 puntos de 2005 a 2008, y las mujeres en la modalidad Telesecundaria, que aumentaron 12 puntos.

Comparación por edad

En los reportes de resultados Excale, de manera habitual, se hace la comparación por edad considerando que una gran cantidad de los estudiantes que se encuentran en situación de extraedad llegaron a ella debido a la repetición de uno o más años. El análisis apunta entonces a proporcionar algunos

5 En éste y otros apartados, el valor puede ser ligeramente diferente al que se obtiene al hacer las operaciones con los datos de la gráfica. Esto es debido a que el dato de la gráfica está redondeado.

6 En la tabla B del Anexo 1 se pueden consultar los errores estándar de los datos mencionados en este párrafo, así como las diferencias significativas y no significativas entre ellos.

7 Al igual que en otros casos, en la figura 7 los datos están redondeados a la unidad. Las diferencias en porcentajes de este párrafo se calcularon utilizando la tabla C del Anexo 1 que tiene cantidades redondeadas a un decimal.

Figura
7

Porcentaje de estudiantes por nivel de logro educativo en Español según sexo

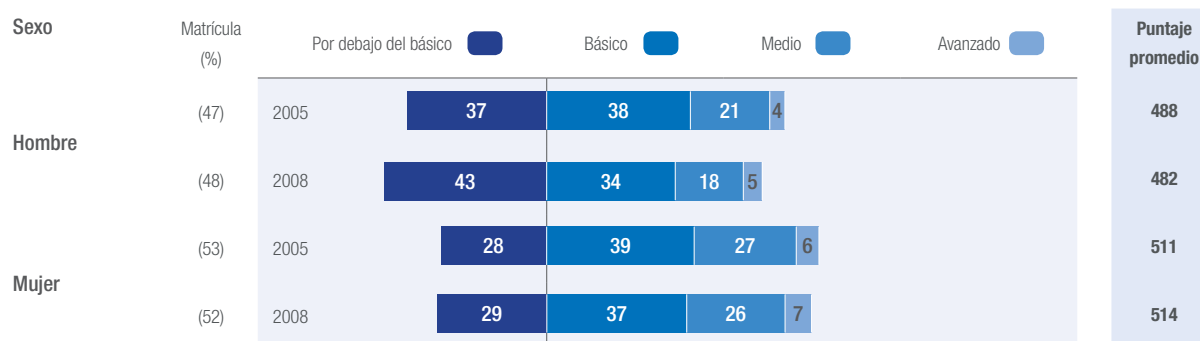


Tabla
XII

Puntuaciones medias en Español por modalidad y sexo

Modalidad educativa	Hombre						Mujer					
	2005		2008		Diferencia*		2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)
Nacional	488	(1.4)	482	(2.6)	-7	(3.0)	511	(1.4)	514	(2.1)	2	(2.6)
General	495	(2.9)	486	(5.3)	-8	(6.0)	514	(2.9)	513	(4.1)	-1	(5.2)
Técnica	490	(2.4)	484	(2.8)	-6	(3.4)	513	(2.1)	518	(2.8)	5	(3.5)
Telesecundaria	440	(3.0)	442	(3.3)	2	(4.2)	465	(2.7)	476	(3.6)	12	(4.3)
Privada	578	(3.4)	554	(6.5)	-25	(7.2)	600	(3.1)	604	(4.2)	3	(5.3)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

indicios sobre si la repetición de cursos permite que los estudiantes en situación de extraedad nivelen sus conocimientos y habilidades con los de los estudiantes que se encuentran en la edad esperada.

La figura 8 muestra las puntuaciones medias y sus respectivos intervalos de confianza para los estudiantes de edad normativa y extraedad en las aplicaciones de 2005 y 2008. Como puede observarse, las puntuaciones medias por grupo variaron muy poco entre 2005 y 2008, los cambios entre años no fueron significativos al interior de cada grupo. Por otro lado, había diferencias significativas en 2005 entre los estudiantes en edad normativa y los estudiantes en situación de extraedad, de 60 puntos, las cuales se conservaron en 2008, con una distancia de 58 puntos⁸.

⁸ La tabla D del Anexo 1 muestra las puntuaciones medias en Español por edad, así como las diferencias entre puntuaciones del mismo grupo para diferente año, y del mismo año para diferente grupo, indicando con negritas las que son significativas.

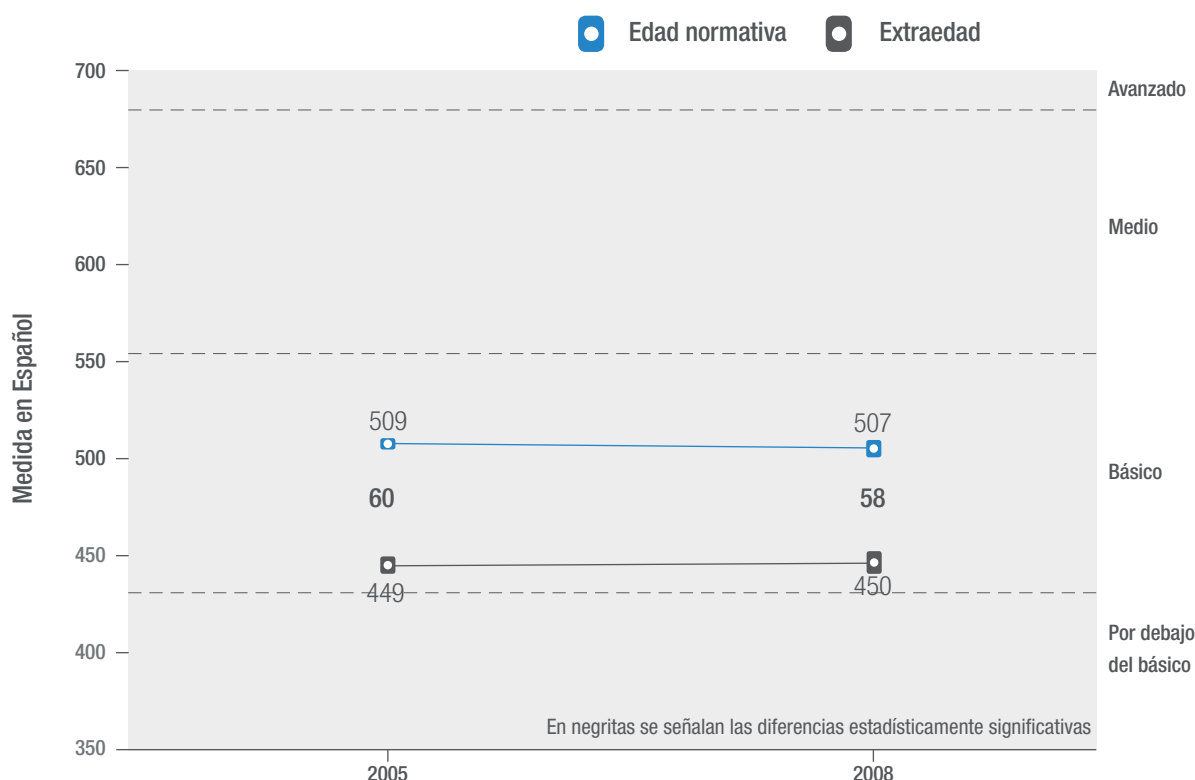
Comparación por edad y modalidad

Al desagregar los grupos de estudiantes en edad normativa y en situación de extraedad por modalidad educativa, sí encontramos algunos grupos con diferencias significativas entre 2005 y 2008. La tabla XIII muestra las puntuaciones medias tanto en 2005 como en 2008 para los alumnos en edad normativa y en extraedad por modalidad educativa.

Se puede observar que, además de las diferencias significativas en Telesecundaria entre 2005 y 2008 ya mencionada para la población en general, al desglosarla por edad esta diferencia se localiza más claramente en el grupo de edad normativa, aunque debe notarse que la diferencia entre años de este grupo y la del grupo en situación de extraedad no son muy distantes (ocho y seis puntos, respectivamente) y que el segundo tiene errores estándar mayores, lo que hace más riguroso su criterio de significancia. También hay diferencias

Figura 8

Comparación de las puntuaciones medias en Español por edad



significativas en la modalidad Privada para el grupo en edad normativa (disminuyó 11 puntos en 2008), y en la modalidad General para el grupo en situación de extraedad (disminuyó 13 puntos).

En suma, encontramos en las pruebas Excale de tercero de secundaria, Español, un aumento significativo de ocho puntos entre 2005 y 2008 en las telesecundarias, pero no en las demás modalidades educativas.

Aunque no resulta estadísticamente significativa, encontramos una disminución de 11 puntos en el mismo periodo en la modalidad Privada. Como consecuencia de estos movimientos, las brechas entre la modalidad con puntuación más baja (Telesecundaria) y el promedio nacional, y también entre Telesecundaria y la modalidad de puntuación promedio más alto (Privada), disminuyeron entre 2005 y 2008.

A nivel nacional, una tercera parte de los estudiantes permanece en el nivel por debajo del básico, y en el caso de Telesecundarias, este valor permanece en los dos años revisados en la mitad de los

estudiantes. Además, las diferencias en la media entre hombres y mujeres, que ya eran significativas en 2005, aumentaron ligeramente debido a que los hombres tuvieron una puntuación significativamente menor en 2008 que la que tuvieron en promedio en 2005. Tomando en cuenta simultáneamente sexo y modalidad, encontramos que los cambios significativos se concentran en un aumento de 2005 a 2008 en mujeres de Telesecundaria, y una disminución en el mismo periodo en hombres de modalidad Privada; esta última, al ser grande (una cuarta parte de una desviación estándar) parece impactar junto con disminuciones no significativas en las modalidades General y Técnica el promedio nacional de los hombres, de tal modo que éste muestra una disminución significativa de siete puntos.

Con respecto a las diferencias por edad, la enorme diferencia en promedio que existía desde 2005 entre los estudiantes en edad normativa y los que se encuentran en situación de extraedad permanece sin cambios en 2008. Al desagregar la información simultáneamente por edad y modalidad encontramos que las diferencias significativas



Tabla
XIII

Puntuaciones medias en Español por modalidad educativa y edad

Modalidad educativa	Edad normativa						Extraedad					
	2005		2008		Diferencia*		2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)
Nacional	509	(1.4)	507	(2.1)	-1	(2.5)	449	(2.3)	450	(3.0)	0	(3.5)
General	511	(2.8)	508	(4.2)	-3	(5.1)	461	(4.2)	447	(5.4)	-13	(6.5)
Técnica	509	(1.9)	510	(2.2)	1	(2.9)	444	(4.2)	451	(5.4)	7	(6.7)
Telesecundaria	461	(2.7)	470	(2.7)	8	(3.5)	423	(3.4)	430	(3.9)	6	(4.6)
Privada	592	(2.7)	582	(4.7)	-11	(5.3)	561	(8.7)	560	(11.2)	-5	(16.1)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

entre 2005 y 2008 se concentran en los alumnos de edad normativa de Telesecundaria (aumento) y Privada (disminución), así como en los alumnos en extraedad de modalidad General (disminución).

Cabría preguntarse: ¿sucedió algo en las aulas de Telesecundaria en estos años que llevara a una mejora significativa, aunque pequeña? ¿Sucedió algo en las aulas de educación privada? Parte de la explicación de la disminución en el promedio en escuelas de sostenimiento privado podría estar relacionada con el aumento de 14%

de planteles escolares de esta modalidad entre 2005 y 2008, si los nuevos planteles no tuvieran la misma calidad en la oferta que los anteriores. Tómese en cuenta que en 2008 el número de planteles de Telesecundaria aumentó de manera similar (en valores absolutos, porque en términos relativos correspondería a un aumento de 3% con respecto a 2005) a la modalidad Privada, obteniendo resultados muy distintos. Por otro lado, ¿parte de lo observado puede ser debido a cambios en la conformación social de los estudiantes en estos grupos?

Capítulo 3.

Resultados y comparación en Matemáticas

Definición de la escala de Matemáticas¹

Al igual que en el caso del Español, la primera prueba Excale para tercero de secundaria fue aplicada en 2005. El diseño general del Excale de Matemáticas tiene como marco de referencia los propósitos, los contenidos y el enfoque señalados en el Plan y programas de estudio 1993, aquél que estudiaron los alumnos de tercero de secundaria evaluados en las dos aplicaciones (2005 y 2008) que se han hecho del Excale 09.

El currículo de Matemáticas Secundaria de 1993 se organiza en tres programas de estudio, uno para cada grado escolar. Los programas de estudio se organizan en cinco áreas: Aritmética, Álgebra, Geometría, Presentación y tratamiento de la información y Nociones de Probabilidad.

El propósito central de estos programas es que los alumnos aprendan a utilizar las Matemáticas para resolver problemas; no solamente los que se resuelven con los procedimientos y técnicas aprendidas en la escuela, sino aquéllos cuyo descubrimiento y resolución requieren de la curiosidad y la imaginación creativa (SEP, 1993a).

En la resolución de un problema matemático es común que se vean implicados de manera simultánea diversos conocimientos, habilidades y destrezas. En este escenario, en términos de evaluación se consideró conveniente desagregarlos con la intención de evaluar en cada reac-

tivo una habilidad o conocimiento a la vez y así poder detectar y comunicar con mayor detalle los aspectos puntuales de un mismo contenido curricular que han sido aprendidos o no por los estudiantes.

Rediseño del Excale de Matemáticas

A partir de la experiencia de 2005, y con la intención de ofrecer una mejor prueba, se rediseñó el Excale 09 de Matemáticas incluyendo 60 nuevos reactivos encaminados a describir con mayor detalle los conocimientos y habilidades de los estudiantes que se ubican en los niveles de dominio más bajo. Este rediseño fue considerado cuidadosamente al hacer los análisis comparativos bajo el método descrito en el capítulo 1.

Para tener una idea general de la estructura del Excale 09 de Matemáticas, en la tabla XIV se presenta la distribución de reactivos de acuerdo con las áreas en que se organiza el currículo de Matemáticas de 1993. En los reportes correspondientes (Backhoff *et al.*, 2006; y Sánchez y Andrade, 2009) se presentan descripciones detalladas de estas pruebas.

¹ Este apartado es una síntesis de la sección correspondiente en Sánchez y Andrade (2009), cuya autoría corresponde a Juan Carlos Xique Anaya y Miguel Ángel León Hernández.

Estructura del Excale 09 de Matemáticas

Áreas del currículo	Habilidades y conocimientos	Reactivos
Aritmética 44.8%	Los números naturales y sus operaciones	29
	Los decimales y sus operaciones	21
	Fracciones	18
	Números con signo	4
	Proporcionalidad	10
Geometría 26.8%	Forma	20
	Sólidos	14
	Medición y cálculo geométrico	15
Álgebra 21.3%	Preálgebra	6
	Monomios y polinomios	9
	Ecuaciones	15
	Funciones	9
Presentación y tratamiento de la información 3.8%	Tablas y gráficas	7
Probabilidad 3.3%	Probabilidad	6
Total		183

La distribución de los reactivos del Excale en cada una de las áreas corresponde de manera aproximada a la carga de contenidos considerados para cada una de las áreas por el currículo nacional de Matemáticas.

Una manera habitual de presentar los resultados de los Excale es a partir de niveles de logro, en los cuales se distribuye la población evaluada. Los cuatro niveles de logro del Excale 09 de Matemáticas se presentan en la tabla xv.

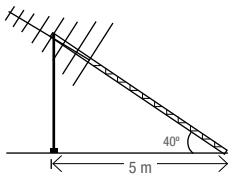
La figura 9 ejemplifica la escala de evaluación del Excale 09². En ella podemos ver los resultados de algunos de los reactivos de la prueba organizados de acuerdo con su grado de dificultad. En la parte superior de este gráfico aparecen los contenidos con mayor grado de dificultad para los estudian-

tes y de manera consecutiva y descendente los reactivos con menor grado de dificultad. En este mapa también se señalan los cuatro niveles de logro educativo y los puntos de corte que los delimitan. Al lado derecho de la descripción del contenido de cada reactivo se informa el porcentaje de respuestas correctas que se obtuvo al resolverlo. Por ejemplo, el reactivo referido a Identificar el radio de un círculo obtuvo 361 puntos de dificultad, con 83% de aciertos y se ubica en el nivel de logro Por debajo del básico.

Resultados comparativos en Matemáticas

Al igual que en el apartado anterior, es importante mencionar que los resultados descritos de manera extensa para cada una de las dos aplicaciones utilizadas en este estudio se presentan en los reportes detallados realizados por el INEE en 2006 y 2009.

² Se muestra la más reciente, la de la aplicación 2008.

Nivel de logro	Reactivos ejemplo
Avanzado Los estudiantes que se ubican en este nivel resuelven problemas que implican potenciación y radicación con números naturales o decimales. Resuelven problemas de reparto proporcional y de proporcionalidad inversa. Dividen y factorizan polinomios. Modelan problemas mediante ecuaciones de segundo grado. Establecen relaciones entre todo tipo de representaciones: tabulares, gráficas y algebraicas de una función lineal o cuadrática. También resuelven problemas que implican el cálculo del área del círculo, así como del área lateral y volumen de cuerpos geométricos; utilizan propiedades o teoremas sencillos para resolver problemas geométricos o de medición (por ejemplo, las razones trigonométricas y el teorema de Pitágoras). Reconocen y realizan transformaciones o movimientos en el plano (simetría, rotación, traslación) en la resolución de problemas de construcción, de medida o de escala. Inferen información a partir de datos presentados en una tabla o en una gráfica. Finalmente, resuelven problemas de probabilidad aplicando la regla de la suma o del producto. Los estudiantes de este nivel dominan también lo que se enuncia en los niveles Medio, Básico y Por debajo del básico.	Un tubo que sostiene una antena se fija perpendicularmente sobre una superficie horizontal por medio de un cable. La distancia de la base del poste a donde se ancla el cable es de 5 m, y el ángulo de inclinación del cable es de 40°. ¿Cuál es la altura del tubo?  $\begin{aligned}\text{sen } 40^\circ &= 0.6427 \\ \text{cos } 40^\circ &= 0.7660 \\ \text{tan } 40^\circ &= 0.8390\end{aligned}$ ☐ 5.9587 m ☐ 3.2139 m ☐ 6.5270 m ☐ 4.1950 m
Medio Los estudiantes que se ubican en este nivel resuelven problemas en los que se utilizan más de dos operaciones con números decimales o fraccionarios, así como la raíz cuadrada con números naturales. Establecen relaciones de orden con números racionales. Resuelven problemas de proporcionalidad con valor unitario no entero o de reparto proporcional, incluyendo porcentajes. Resuelven problemas que implican usar la jerarquía de operaciones. Asimismo, multiplican polinomios y modelan situaciones mediante una función lineal o cuadrática. Modelan también situaciones que implican sistemas de ecuaciones con dos incógnitas. Establecen relaciones entre dos cualesquiera de las formas de representación de una función lineal o cuadrática. Establecen relaciones entre una tabla de valores y su expresión algebraica. Resuelven ecuaciones de segundo grado. Reconocen procedimientos que llevan a la construcción de una figura geométrica. Utilizan las propiedades de lugares geométricos (alturas, bisectrices, mediatrices) en la resolución de problemas de construcción, escala o medición. Asimismo, resuelven problemas que implican el cálculo del perímetro del círculo, utilizan las propiedades de lugares geométricos (alturas, bisectrices, mediatrices) en la resolución de problemas de construcción de medidas o escalas. Finalmente, interpretan información contenida en tablas o gráficas de distintos tipos. Resuelven problemas que implican el uso de medidas de tendencia central. Resuelven problemas de conteo y determinan la probabilidad de eventos mutuamente excluyentes. Los estudiantes de este nivel dominan también lo que se enuncia en los niveles Básico y Por debajo del básico.	¿En cuál de las siguientes mezclas el agua tiene mayor sabor de uva? ☐ 4 vasos de jugo por 2 lts de agua ☐ 5 vasos de jugo por 15 lts de agua ☐ 2 vasos de jugo por 5 lts de agua ☐ 3 vasos de jugo por 7 lts de agua

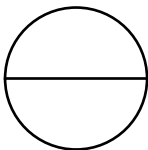
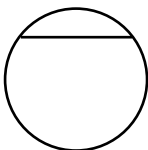
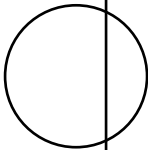
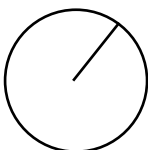
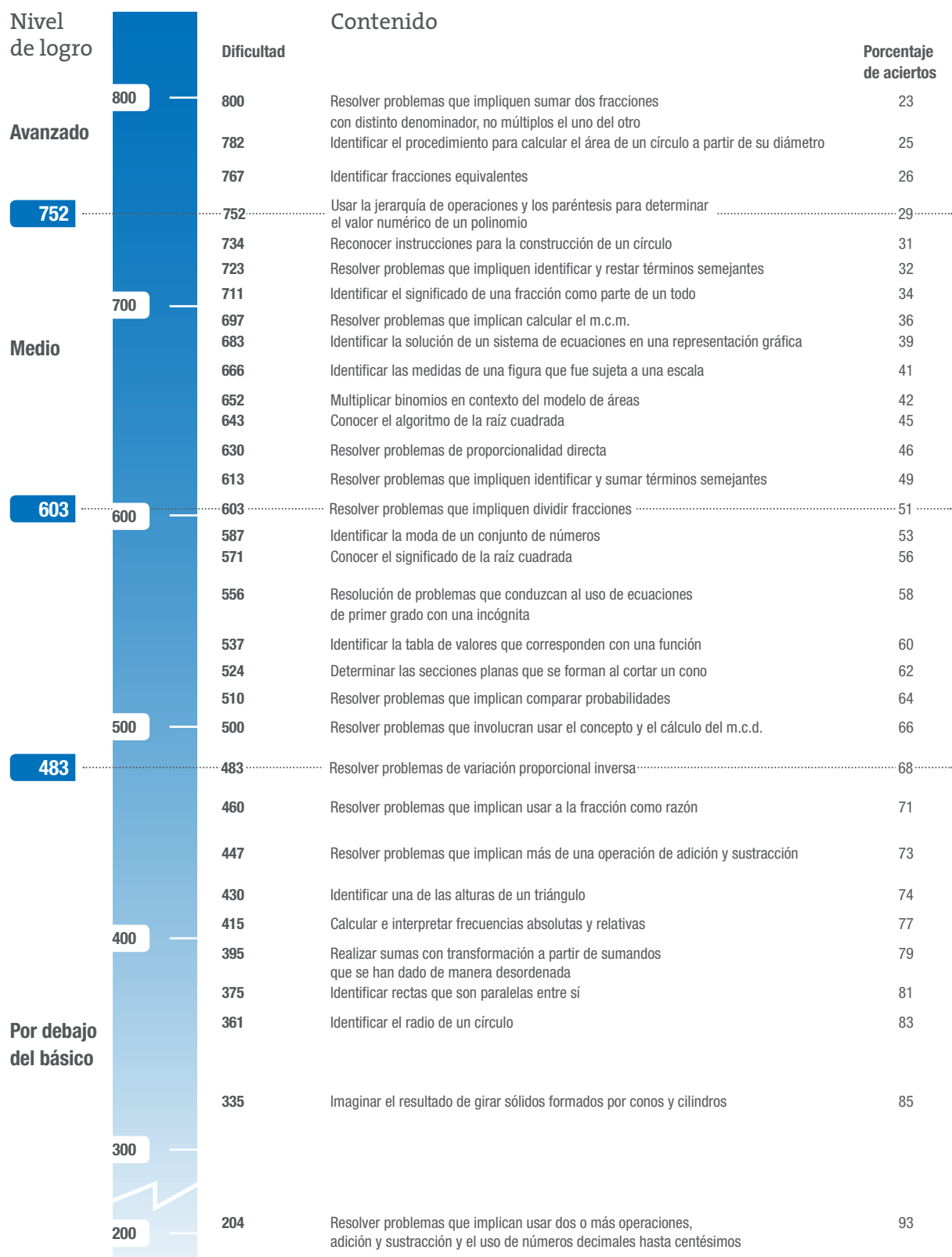
Nivel de logro	Reactivos ejemplo
Básico Los estudiantes que se ubican en este nivel de logro resuelven problemas aditivos multiplicativos y de potenciación que impliquen dos o más operaciones con números naturales y enteros. Resuelven problemas mediante el uso del mínimo común múltiplo y máximo común divisor. Establecen relación de orden entre números racionales positivos. Resuelven problemas de proporcionalidad con valor unitario entero. Identifican situaciones de proporcionalidad inversa. Suman y restan polinomios. Traducen del lenguaje verbal a una expresión de la forma $ax+b=c$ y resuelven ecuaciones de primer grado con una incógnita, así como sistemas de ecuaciones con dos incógnitas. Igualmente utilizan las propiedades de las figuras (por ejemplo, ángulos, lados y diagonales) en la resolución de problemas de medición. También calculan el perímetro y el área de figuras básicas (triángulos, cuadriláteros y polígonos regulares) y el volumen de cuerpos geométricos. Identifican el procedimiento que permite calcular la mediana. Leen información contenida en tablas o gráficas de frecuencia. Resuelven problemas de conteo y determinan la probabilidad de eventos simples. Los estudiantes de este nivel dominan también lo que se enuncia en el nivel Por debajo del básico.	En promedio don José cosechó 100 naranjas por cada uno de los 20 árboles que tiene, y las empacó en costales con 25 naranjas cada uno. ¿Cuántos costales llenos de naranja tiene? ☐ 80 ☐ 125 ☐ 2000 ☐ 50 000
Por debajo del básico Los estudiantes que se ubican en este nivel leen y escriben números naturales y establecen relaciones de orden entre ellos; resuelven problemas aditivos que impliquen una sola operación con números naturales, enteros, decimales o fraccionarios, así como problemas multiplicativos con números naturales. Asimismo, identifican situaciones de proporcionalidad directa y establecen relaciones entre una tabla de valores y su gráfica en funciones lineales o cuadráticas. Adicionalmente, identifican figuras o cuerpos geométricos a partir de sus elementos o características. Calculan el perímetro y el área de figuras básicas (triángulos, cuadriláteros y polígonos regulares). Identifican la moda y la media en un conjunto de datos. Finalmente, estiman y comparan la probabilidad de eventos simples.	¿En cuál de los siguientes círculos se ha trazado sólo uno de sus radios? ☐  ☐  ☐  ☐ 

Figura
9

Mapa parcial de reactivos: Matemáticas³



³ Si el lector desea conocer la información de la totalidad de los reactivos que conformaron esta prueba, puede consultar el Anexo 3 de Sánchez y Andrade (2009).

Puntuaciones medias en Matemáticas por modalidad educativa

Modalidad educativa	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)		(EE)
Nacional	500	(1.4)	504	(2.1)	4	(2.4)
General	500	(2.8)	506	(3.5)	6	(4.3)
Técnica	497	(1.8)	499	(2.8)	1	(3.4)
Telesecundaria	473	(2.2)	480	(3.0)	7	(3.7)
Privada	574	(3.3)	579	(5.5)	5	(6.2)

* Las diferencias pueden no ser exactas debido al redondeo.

Comparación por modalidad

Como se ha comentado en el capítulo anterior, además de las diferencias en los resultados entre diferentes años de aplicación a nivel nacional, es importante dar seguimiento a las diferencias en los resultados entre modalidades educativas, porque de este modo se pueden analizar con más detalle los cambios (o no) en el transcurso del tiempo.

En la tabla XVI se presentan las puntuaciones medias obtenidas en 2005 y 2008 por la población nacional y por las modalidades en estudio.

No se encontraron diferencias significativas entre las medias de 2005 y 2008 ni a nivel nacional ni en las diferentes modalidades. Adicionalmente, puede observarse que el ordenamiento entre modalidades es idéntico en ambos años.

La figura 10 muestra estos mismos datos graficados. Al igual que en otras gráficas de este reporte, la media se señala con la figura en blanco dentro de cada rectángulo. Los límites inferior y superior de cada rectángulo indican el intervalo de confianza. Además, se señalan los niveles de logro en el margen derecho y sus respectivos puntos de corte con líneas punteadas tenues que atraviesan horizontalmente la gráfica. Como puede observarse, aunque no se encontraron diferencias significativas, para todos los grupos las puntuaciones medias aumentaron ligeramente en 2008. Cabe destacar que, para el caso de la modalidad Técnica, el pequeño aumento supone acercarse al límite inferior del nivel Básico, y para la población nacional y la modalidad General, empezar a despegarse del mismo. Para la modalidad Privada, el

ligero aumento implica acercarse al límite inferior del nivel Medio.

Con respecto a la distribución de los alumnos en los niveles de logro, la Figura 11 muestra el porcentaje de estudiantes que se encontraban en los diferentes niveles de logro en Matemáticas, tanto en 2005 como en 2008.

Puede observarse que en todos los casos el nivel Básico disminuye en 2008 con respecto a 2005. A nivel nacional, esta disminución de 2% se corresponde con pequeños aumentos en los niveles Por debajo del básico, Medio y Avanzado⁴.

En el caso de la modalidad Telesecundaria, la disminución en el nivel Básico es de 2.6% y se corresponde con aumentos en los niveles Medio y Avanzado de 2.1 y 0.4%, respectivamente.

En la modalidad Técnica, la disminución de 2.7% en el nivel Básico está acompañada de una ligera disminución en el nivel Medio y éstas se corresponden con aumentos de 2 y 0.8%, respectivamente, en los niveles Por debajo del básico y Avanzado.

Para la modalidad General, la disminución de 1.5% en el nivel Básico se corresponde con aumentos de 1 y 0.5% en los niveles Medio y Avanzado, respectivamente.

Finalmente, para la modalidad Privada, la disminución en el nivel Básico de 1.7% se acompaña

⁴ Los valores con decimales a partir de los cuales se hicieron los cálculos de diferencias entre porcentajes a nivel nacional y por modalidades se encuentran en la tabla F del Anexo 2.



de otra disminución en el nivel Medio, de 1.9%, que se corresponden con aumentos de 0.8 y 2.7% en los niveles Por debajo del básico y Avanzado, respectivamente.

Estas diferencias sólo son indicios, ya que no tienen significancia estadística. Aun así, debe mencionarse que pequeños cambios sostenidos —en que las disminuciones en los niveles de logro Por debajo del básico y Básico se acompañen de aumentos en los otros dos niveles (como es el caso de las moda-

lidades General y Telesecundaria) con el tiempo— llevarían a cambios significativos y positivos en las puntuaciones de las diferentes modalidades.

Brechas entre modalidades

Como se mencionaba en el capítulo anterior, es de interés conocer si las distancias entre modalidades han aumentado o disminuido.

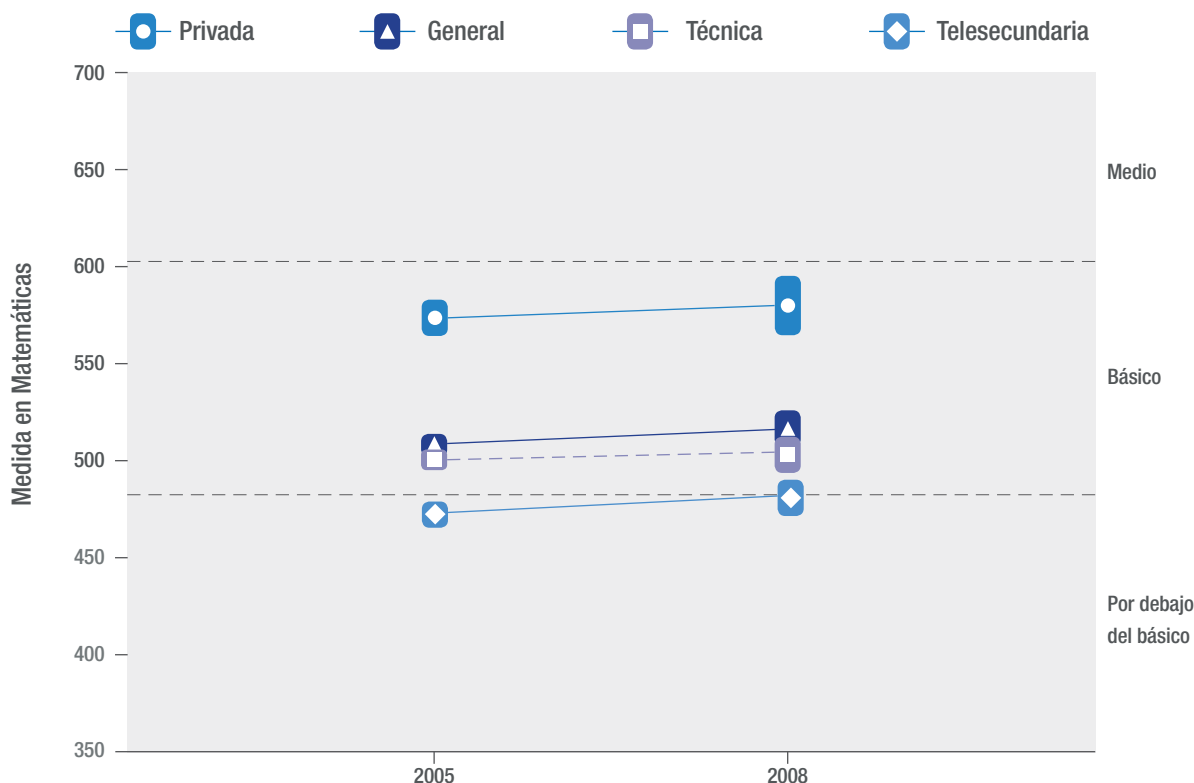
Al considerar nuevamente los datos de la tabla XVI puede observarse que la diferencia en la puntuación media entre la Telesecundaria (modalidad con puntuación más baja) y la modalidad Privada (puntuación más alta) es de 101 puntos en 2005 y de 99 en 2008.

La diferencia en la puntuación media entre la Telesecundaria y la media nacional era en 2005 de 27 puntos y en 2008 de 24 puntos.

Mientras que en Español la disminución de la brecha se dio debido a que la modalidad con puntuación alta decreció de 2005 a 2008, en Matemáticas la modalidad con puntuación alta aumentó ligeramente, pero la modalidad con

Figura 10

Comparación de las puntuaciones medias en Matemáticas por modalidad educativa



puntuación baja aumentó más en el mismo lapso. Ésta parece una manera deseable de cerrar las brechas: todos los grupos progresan, pero los que tienen menores puntajes progresan en mayor medida, o a mayor velocidad.

Aunque estos resultados parecen ir en la dirección correcta (disminución de las brechas), la di-

ferencia entre años es muy pequeña, de modo que debe mencionarse sólo como un indicio de mejora.

Dispersión dentro de las modalidades

La dispersión a nivel nacional y por modalidad educativa nos indica las distancias que hay en-

Figura 11

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas por modalidad educativa

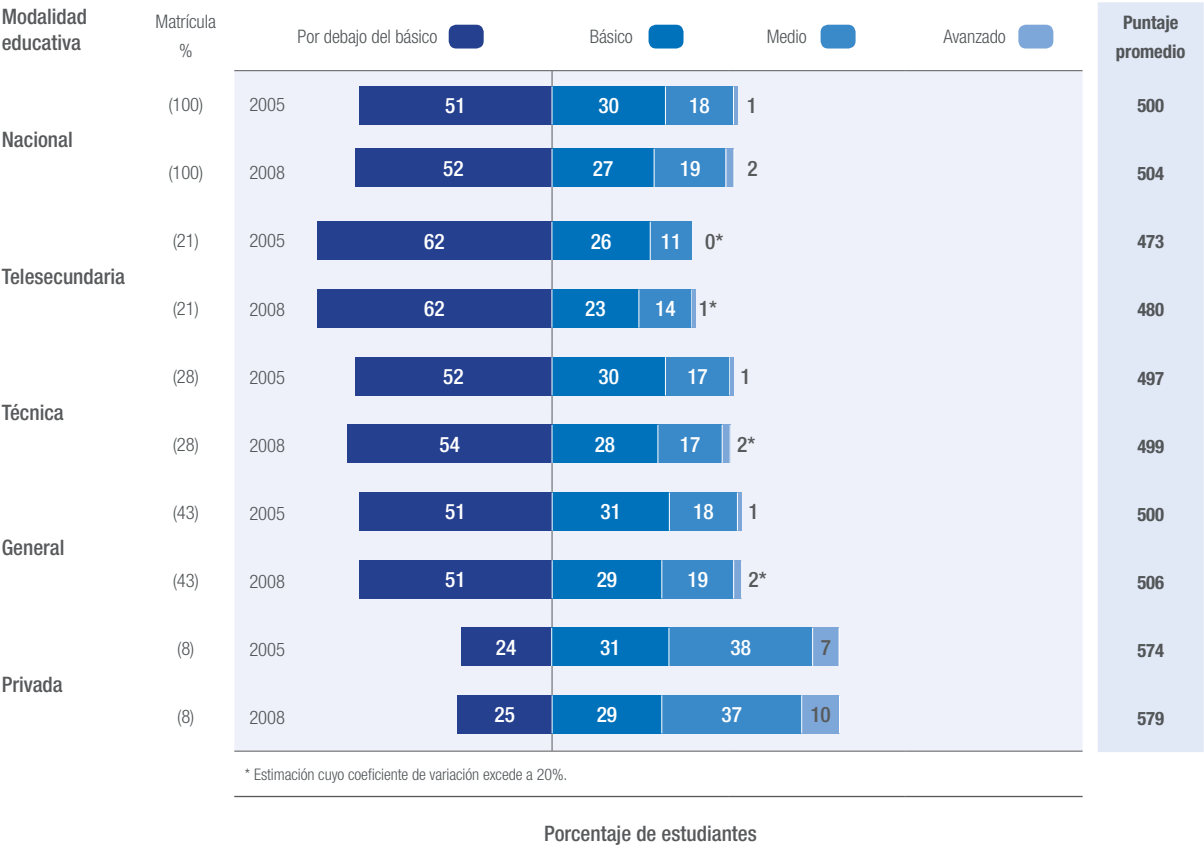


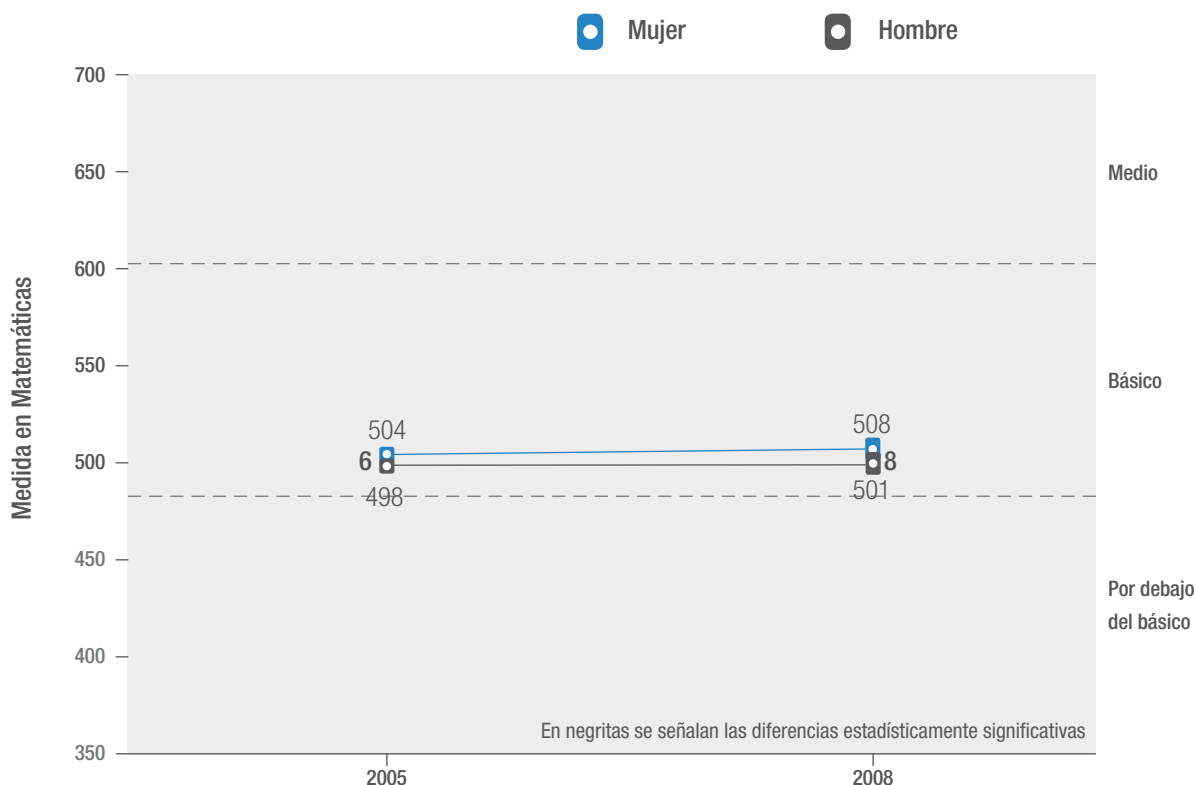
Tabla XVII

Dispersión de puntajes en Matemáticas, por modalidad

Modalidad educativa	2005				2008			
	D.E.	(EE)	C.V.	(EE)	D.E.	(EE)	C.V.	(EE)
Nacional	100	(0.9)	20	(0.2)	101	(1.3)	20	(0.2)
General	97	(1.8)	19	(0.4)	96	(2.1)	19	(0.4)
Técnica	95	(1.0)	19	(0.2)	98	(1.8)	20	(0.3)
Telesecundaria	94	(1.6)	20	(0.3)	96	(1.9)	20	(0.4)
Privada	105	(1.7)	18	(0.3)	111	(3.1)	19	(0.5)

Figura 12

Comparación de las puntuaciones medias en Matemáticas por sexo



tre los alumnos al interior de cada uno de estos grupos. En la tabla XVII se muestran valores de dispersión de puntajes por modalidad. La desviación estándar se acompaña del coeficiente de variación por ser éste una expresión de homogeneidad relacionada con la media aritmética del grupo de referencia. Como puede observarse, tanto la desviación estándar como el coeficiente de variación nos muestran que no hay diferencias al interior de las distintas modalidades entre 2005 y 2008 con respecto a la dispersión de sus puntajes.

Comparación por sexo

La intención principal de comparar los resultados de los Excale por sexo es la de tener elementos sobre equidad en la educación. Sería de esperar que no haya brechas entre los resultados de los dos sexos, o que éstas sean pequeñas y similares a las que muestran repetidamente estudios de diferentes partes del mundo (por ejemplo, pequeñas diferencias en el aprendizaje de la lengua a favor de las mujeres).

La figura 12 muestra las puntuaciones promedio en Matemáticas para mujeres y hombres en 2005 y 2008.

Puede observarse que la diferencia de puntuación entre mujeres y hombres tiene la misma tendencia en los dos años estudiados: pequeñas diferencias significativas a favor de los hombres⁵. La distancia de estas diferencias es prácticamente la misma: seis puntos en 2005 y ocho en 2008.

En la figura 13 puede observarse la distribución de los resultados en los diferentes niveles de logro por sexo.

Tanto para mujeres como para hombres, el nivel de logro Básico tuvo ligeras disminuciones en

⁵ En los reportes anuales de los Excale 2005 y 2008 se había indicado erróneamente que las diferencias en Matemáticas de un mismo año entre hombres y mujeres no eran significativas. La tabla G del Anexo 2 presenta las puntuaciones medias en Matemáticas por sexo, con un decimal y sus errores estándar. La manera correcta de hacer la comparación de medias en este caso es la sugerida por la OCDE en PISA Data Analysis Manual (2009).

Figura
13

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y sexo

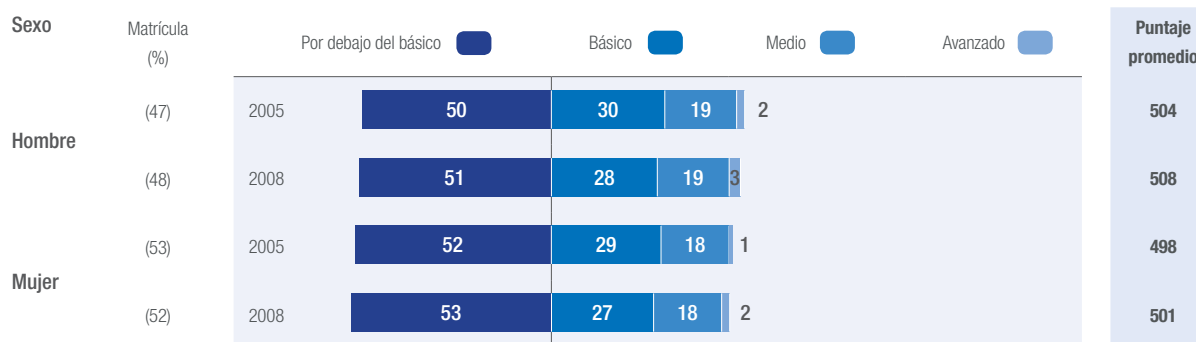


Tabla
XVIII

Puntuaciones medias en Matemáticas por modalidad y sexo

Modalidad educativa	Hombre						Mujer					
	2005		2008		Diferencia*		2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)
Nacional	504	(1.5)	508	(2.6)	5	(2.9)	498	(1.7)	501	(2.4)	3	(2.8)
General	505	(3.0)	511	(4.5)	6	(5.0)	497	(3.1)	502	(4.2)	5	(5.5)
Técnica	501	(2.2)	502	(3.3)	1	(4.1)	495	(2.0)	496	(3.2)	0	(3.7)
Telesecundaria	475	(2.7)	481	(4.1)	6	(5.1)	472	(2.6)	479	(3.8)	7	(4.5)
Privada	578	(4.0)	584	(5.9)	6	(7.0)	573	(3.7)	576	(6.3)	3	(7.1)

* Las diferencias pueden parecer inexactas debido a que la tabla utiliza valores redondeados de las medias, sin decimales, que sí se tomaron en cuenta para el cálculo de las diferencias.

2008, de 2 y 2.4%, respectivamente. También en ambos casos aumentó el porcentaje de estudiantes en el nivel Por debajo del básico (mujeres 1.4%; hombres 0.9%), en el nivel Medio (0.4% cada uno de los dos grupos) y en el nivel Avanzado (0.2% las mujeres y 1.1% los hombres).

Comparación por sexo y modalidad

La tabla XVIII nos muestra los resultados para 2005 y 2008 simultáneamente por modalidad educativa y sexo.

Al igual que en las desagregaciones sólo por sexo, encontramos que en este desglose no hay diferencias significativas entre los dos años estudiados en ninguno de los grupos.

Un dato interesante es que, aunque las diferencias no sean significativas, en todos los casos la

puntuación media de 2005 es menor que la de 2008, tanto en mujeres como en hombres, en todas las modalidades.

Comparación por edad

Se ha mencionado en el capítulo anterior que, dada la relación que existe entre la situación de extraedad y la repetición de cursos, se considera a la comparación entre las puntuaciones de los estudiantes en edad normativa y en situación de extraedad como una manera de verificar si los estudiantes, al repetir cursos, nivelan sus conocimientos y habilidades con respecto a los alumnos que no los han repetido.

La figura 14 muestra las puntuaciones promedio de la población en edad normativa y en situación de extraedad tanto de 2005 como de 2008. Como puede observarse, los ligeros aumentos en la media de ambas poblaciones se corresponden de manera

Figura 14

Comparación de las puntuaciones medias en Matemáticas por edad

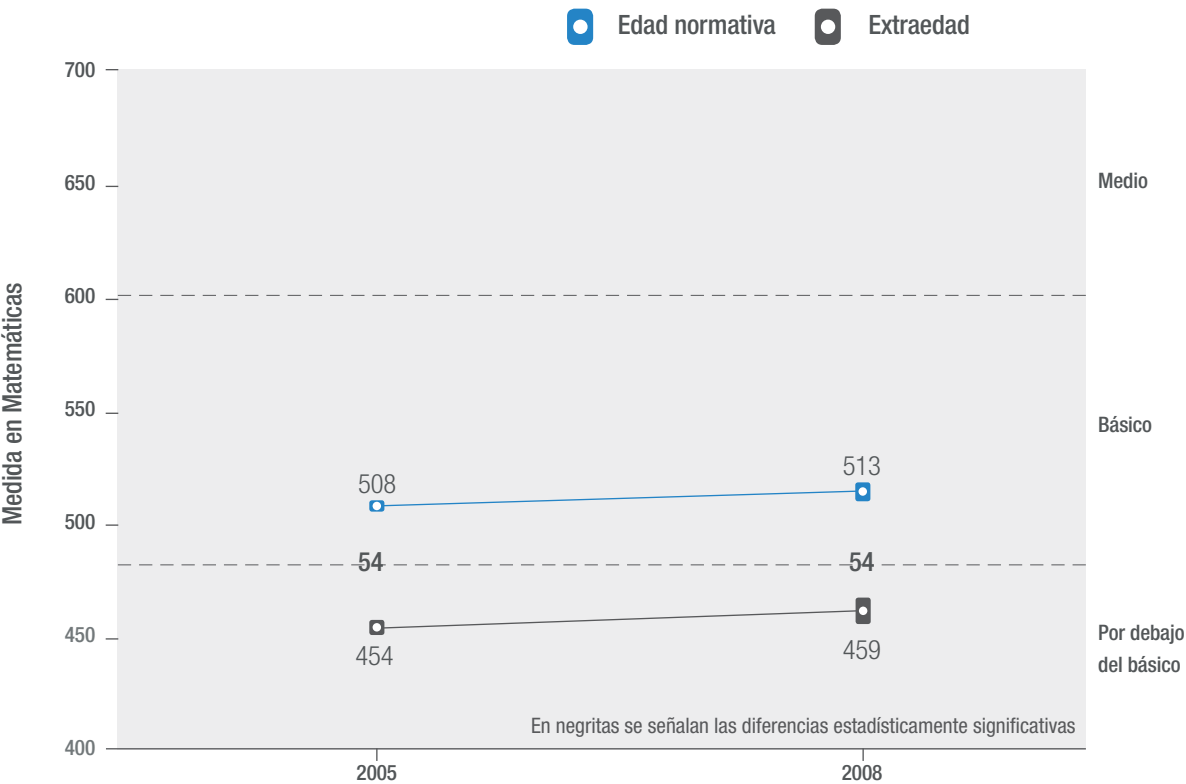
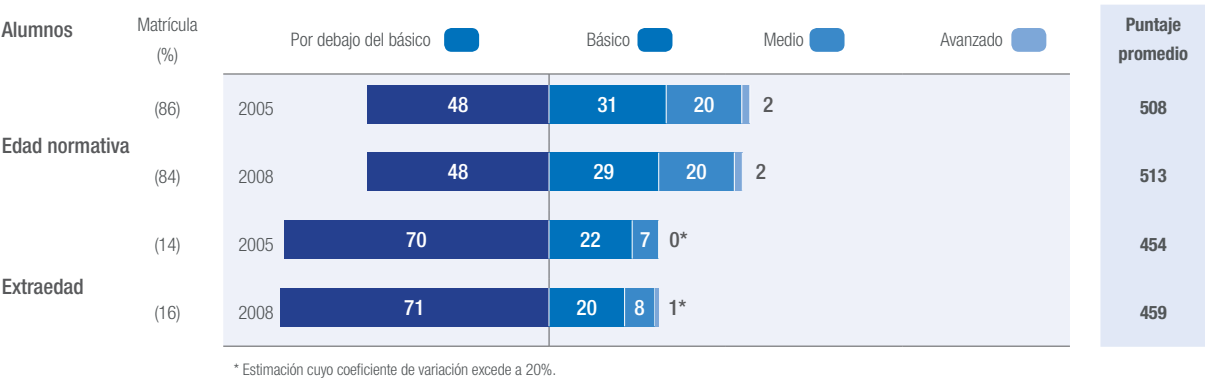


Figura 15

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y edad



exacta, de modo que el tamaño de la brecha entre los dos grupos de edad no cambió en el tiempo.

La figura 15 muestra los resultados por grupo de edad, para ambos años de estudio, desglosados por niveles de logro⁶.

6 La tabla I del Anexo 2 muestra las puntuaciones medias en Matemáticas por grupo de edad y por año de estudio, así como sus errores estándar.

El redondeo de las cifras, expuesto en la gráfica, hace ver en varios casos que no hay cambio. Sin embargo, utilizando los datos de la tabla J del Anexo 2, encontramos que el nivel Básico disminuyó su porcentaje de alumnos en 2008, tanto los de edad normativa (1.8% menos) como los alumnos en situación de extraedad (2.3% menos).

El número de alumnos en el nivel Por debajo del básico aumentó en ambos grupos de edad: 0.4%

Puntuaciones medias en Matemáticas por edad y modalidad

Modalidad educativa	Edad normativa						Extraedad					
	2005		2008		Diferencia*		2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)	Media	(EE)
Nacional	508	(1.5)	513	(2.4)	5	(2.6)	454	(1.9)	459	(2.7)	6	(3.3)
General	507	(2.8)	513	(4.0)	5	(4.8)	454	(4.2)	465	(4.6)	14	(7.2)
Técnica	504	(1.7)	507	(2.9)	3	(3.4)	450	(3.9)	446	(4.4)	-3	(5.4)
Telesecundaria	483	(2.6)	494	(3.0)	11	(4.0)	443	(2.8)	443	(4.6)	0	(5.3)
Privada	578	(3.3)	585	(6.0)	6	(6.8)	545	(8.7)	553	(12.6)	5	(14.0)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

en el grupo de edad normativa y 0.8% en el grupo de extraedad. También en ambos grupos aumentó el porcentaje de alumnos en el nivel Medio (0.7% el grupo en edad normativa y 1% el grupo de extraedad) y en el nivel Avanzado (0.8% el grupo en edad normativa y 0.5% el grupo en extraedad).

Comparación por edad y modalidad

En la tabla XIX pueden verse las puntuaciones medias en Matemáticas por grupo de edad (normativa o extraedad) y modalidad educativa. Se muestran las diferencias entre los dos años de aplicación. Como puede observarse, las diferencias significativas se concentran en los estudiantes de Telesecundaria que se encuentran en edad normativa. Este dato nos permite afinar la percepción que se había desarrollado al analizar la tabla XVII, donde se comparaban las medias de la población sólo por modalidades educativas. En ella, ciertamente se observaba que la diferencia mayor entre años estaba en Telesecundaria, pero ésta no resultaba significativa.

En resumen, al comparar los resultados en Matemáticas en los años 2005 y 2008 podemos afir-



mar que sólo hay un avance significativo, aunque pequeño, en la población que se encuentra en edad normativa y en las telesecundarias. Los demás grupos, tanto por modalidad como por sexo y edad, muestran resultados iguales en Matemáticas entre los dos años de aplicación.

Desagregación por entidades

Debido a que las aplicaciones de 2005 y 2008 se hicieron a muestras representativas a nivel estatal, es posible analizar la información de ambas haciendo desagregados por entidades. Como se ha comentado repetidamente al presentar los informes de las pruebas Excale, hacer una comparación simple de los estadísticos de logro de diferentes entidades, sin tomar en cuenta las diferencias en la composición de las modalidades educativas presentes en cada una de ellas, sería muy injusto. Tómese como ejemplo extremo un par de entidades ficticias, una en la que 90% de la matrícula pertenece a secundarias generales y otra en la que 90% pertenece a telesecundarias; esta simple distribución de los estudiantes generaría claramente puntos de partida distintos que harían injusta cualquier comparación.

Para atenuar el efecto de la distribución de la matrícula y poder presentar resultados estatales de una manera más equilibrada, en los reportes de resultados Excale se han presentado los resultados desagregados simultáneamente por entidad y por modalidad, indicando si los resultados de cada entidad y modalidad se encuentran significativamente por arriba o por debajo de la media nacional de esa modalidad. Debe reconocerse que esta solución es parcial, pues aún subsisten importantes diferencias entre entidades que no son consideradas por el análisis. Dos ejemplos fundamentales son la situación socioeconómica de la población de cada entidad y las diferencias geográficas que en algunos casos permiten el fácil acceso a todas las regiones de la entidad y en otros casos no. Sin embargo, se opta por esta solución parcial para poder dar información útil a nivel estatal.

Con estas mismas precautorias se presentan a continuación dos tablas de síntesis que muestran

las entidades que se encontraron por arriba del promedio nacional de cada modalidad en 2005 y 2008 (tabla XX) y las entidades que se encontraron por debajo del promedio nacional de cada modalidad en los mismos años (tabla XXI).

En este informe comparativo lo que más interesa de dichas tablas es identificar las entidades que repiten su posición entre años, aunque también es informativo observar cuándo una misma entidad en una misma modalidad se repite, a pesar de que sea en diferente asignatura o en diferente año. Estas relaciones no son las únicas ni las más importantes que se podrían establecer para el análisis serio de la situación educativa de una entidad; para ello deberán tomarse en cuenta todos los cruces modalidad-asignatura en que se encuentre dicha entidad, además de considerar que las entidades que no aparecen en ninguna de las dos tablas en algún cruce modalidad-asignatura se encuentran en la media nacional en dicha modalidad-asignatura.

Con respecto a las secundarias generales, ninguna entidad se repite en Español o en Matemáticas durante los dos años de comparación, aunque debe mencionarse que Veracruz resultó por encima de la media en ambas asignaturas en 2005, y San Luis Potosí también en ambas en 2008.

En secundarias técnicas, el Distrito Federal repite por encima de la media en Español, en 2005 y 2008; Guerrero aparece por debajo de la media nacional de la modalidad tanto en Español como en Matemáticas, en ambos años; Michoacán se encuentra por debajo de la media nacional de la modalidad en Matemáticas en los dos años.

Tabla
XX

Entidades por **arriba de la media** nacional por asignatura y por modalidad educativa, 2005 y 2008

Modalidad educativa	Español		Matemáticas	
	2005	2008	2005	2008
General	Veracruz	Querétaro San Luis Potosí Tlaxcala	Veracruz	San Luis Potosí
Técnica	Distrito Federal Estado de México Tlaxcala Veracruz	Chihuahua Distrito Federal	Distrito Federal Estado de México Tlaxcala	Aguascalientes Guanajuato
Telesecundaria	Guanajuato Estado de México Tlaxcala	Guanajuato Tlaxcala Veracruz	Jalisco Estado de México	Guanajuato Veracruz
Privada	Distrito Federal	Distrito Federal Estado de México	Distrito Federal	Coahuila

Tabla
XXI

Entidades por **debajo de la media** nacional por asignatura, por modalidad educativa, 2005 y 2008

Modalidad educativa	Español		Matemáticas	
	2005	2008	2005	2008
General	Coahuila Durango Guerrero Michoacán Oaxaca	Jalisco Nayarit Sonora Yucatán	Coahuila Tabasco	Durango Jalisco Michoacán Yucatán
Técnica	Chiapas Guerrero Michoacán Nayarit Oaxaca Yucatán Zacatecas	Guerrero Quintana Roo Tamaulipas	Coahuila Guerrero Michoacán Nuevo León Sonora Yucatán	Guerrero Michoacán Tabasco
Telesecundaria	Chiapas Michoacán Oaxaca	Chiapas Oaxaca	Chiapas Michoacán Tabasco	Chiapas Guerrero Hidalgo Michoacán Nayarit San Luis Potosí Tabasco
Privada	Baja California Campeche Michoacán Sonora	Baja California Campeche Jalisco Michoacán Veracruz	Campeche Michoacán Sonora	Baja California Michoacán Veracruz Yucatán



En telesecundarias, Guanajuato y Tlaxcala se encuentran por arriba de la media de la modalidad en Español, tanto en 2005 como en 2008; Guanajuato además se encuentra por encima de la media en Matemáticas en 2008. El Estado de México se encuentra por encima de la media en ambas asignaturas en 2005, y Veracruz se encuentra por encima de la media en ambas asignaturas en 2008. Chiapas se encuentra por debajo de la media de la modalidad en ambas asignaturas en los dos años. Oaxaca se encuentra por debajo de la media en Español en los dos años.

Michoacán y Tabasco se encuentran por debajo de la media en Matemáticas en los dos años.

En la modalidad Privada, el Distrito Federal se encuentra por arriba de la media nacional en Español en los dos años y en Matemáticas en 2005. Michoacán se encuentra por debajo de la media en esta modalidad en las dos asignaturas en los dos años. Baja California y Campeche se encuentran por debajo de la media nacional de la modalidad en Español.

Conclusiones

Las pruebas Excale que el INEE aplica a partir de 2005 han tenido desde su diseño inicial la intención de presentar resultados comparativos en el tiempo. Estos resultados deben utilizarse para monitorear el Sistema Educativo Nacional a partir del rendimiento de los estudiantes, como conjunto, al paso del tiempo. Los datos permitirán, en un futuro, analizar las tendencias en el logro de aprendizajes de grandes grupos de alumnos: a nivel nacional, por modalidades educativas, por sexo, por edad. Es necesario recordar que los resultados de las pruebas Excale no deben ser utilizados para evaluar a los alumnos a nivel individual ni a nivel de aula, de escuela o de zona escolar, y que la relación que tienen estos resultados con el desempeño de los profesores o escuelas es indirecto, de modo que tampoco sirven para hacer buenos juicios al respecto.

En este reporte se presentan resultados referentes a las aplicaciones 2005 y 2008 del Excale 09 (tercero de secundaria) de Español y Matemáticas; aún no se reportan tendencias propiamente dichas (dado que éstas requieren al menos tres puntos de comparación), pero sí se ofrecen nuevos elementos de análisis con respecto a los informes previos de estas aplicaciones, pues permiten ver el movimiento o la falta del mismo entre los dos puntos evaluados en el tiempo.

El capítulo 1 reseñó con cierto detalle el método utilizado para comparar la información obtenida en los dos años de referencia. Esto nos parece relevante, debido a que se establece que los resultados no se comparan *a priori*, como si se refirieran a una misma escala, sino que se reconoce que se refieren a escalas con ciertos elementos en común pero no idénticas, que por lo tanto requieren un cuidadoso análisis para decidir la mejor manera de comparar sus resultados. Se ofrece a la vista pública el método empleado con la intención de

posibilitar su reproducción y su revisión. Además, existe la intención de que los siguientes reportes de las aplicaciones Excale integren en un solo documento la información de la aplicación más reciente y la comparativa de la misma y las aplicaciones anteriores. Para hacer más manejables dichos materiales, se hará en ellos referencia al capítulo 1 de este volumen, para evitar entrar en detalles en cada nueva publicación, en tanto no cambien los procedimientos de comparación referidos.

Para poner en la misma escala los resultados de las pruebas de 2005 y 2008 fue necesario llevar a cabo un proceso de vinculación de dos etapas (generalmente conocido como *linking*) entre ellas, que requirió el uso de una muestra puente adicional a la muestra operativa de 2008. A la muestra puente, con representatividad solamente nacional, se le administraron las pruebas tal cual se usaron en el año 2005. Esta estrategia de comparación fue implementada para evitar los efectos de confusión en los resultados debido a los cam-

bios de las pruebas y así fortalecer la estabilidad en la medida de tendencias.

Dichos efectos de confusión son causados fundamentalmente debido a que el supuesto de la independencia local de los modelos de la Teoría de Respuesta al Ítem (TRI) no siempre se cumple en los datos empíricos. Es decir, la respuesta de los alumnos será afectada cuando hay cambios en las pruebas como el orden de las preguntas, el formato de las pruebas, la forma de administrarlas, etc. Cuando hay cambios, los parámetros de los reactivos asociados a los modelos TRI tienden a ser no robustos, por lo que los estimados de dichos parámetros bajo una configuración de prueba en particular no siempre son apropiados para otra configuración de prueba. Si no son implementadas estrategias adecuadas para evitar dichos efectos de confusión en la medición de tendencias, aumenta el riesgo de hacer inferencias falsas, como detectar avances o retrocesos donde en realidad no los hay, o lo contrario, inferir que no hay avances donde sí existen.

En la primera etapa de vinculación, la escala de medida utilizada ya en 2005 se trasladó a la muestra puente mediante la calibración concurrente de reactivos. Es decir, las respuestas de los alumnos obtenidas en 2005 se unieron a las respuestas de los alumnos de la muestra puente —pues se utilizaron exactamente las mismas pruebas— y se realizó la calibración de reactivos. En la segunda etapa, la escala de la muestra puente se trasladó a la muestra operativa de 2008 utilizando el hecho de que tuvieron la población en común, es decir, la muestra puente y la muestra operativa de 2008 fueron seleccionadas utilizando el mismo diseño muestral. Este arreglo, así como los elementos que sí comparten ambas aplicaciones (mismo marco de referencia, diseñadas para mostrar evidencia sobre el mismo constructo, enfocadas para obtener inferencias sobre la misma población objetivo, resultados de bajo impacto en ambos casos y mismos procedimientos de aplicación) permitieron poner los resultados de ambas aplicaciones en una escala de medida común.

Como valoración general se puede observar que no hay grandes cambios entre los resultados obtenidos al aplicar los Excale de Español y Matemáticas entre 2005 y 2008. Esto se puede considerar favorable a la educación, tomando en cuenta que hubo un aumento en el número de escuelas de

9%, y de matrícula de 4% entre estos años, lo cual significa que se incorporaron más estudiantes (y probablemente de grupos sociales más desfavorecidos, aunque el incremento en la matrícula fue proporcional en las diferentes modalidades mencionadas en el presente reporte, de tal modo que su distribución no cambió en este periodo) sin que disminuyeran los resultados educativos. Las mejoras significativas en el logro entre años se concentran en la Telesecundaria para ambas asignaturas, pero aun éstas son muy pequeñas. ¿Sería de esperar cambios en los resultados en las distintas modalidades educativas? Sólo si hubiera acciones de mejora en el periodo 2005–2008 que parecieran claramente relacionadas con el desempeño de los estudiantes. Al parecer, no las hubo. ¿Qué sucedió en la Telesecundaria, en la que sí hay mejoras significativas en el logro?

Si queremos que los estudiantes se preparen mejor, que tengan mejores oportunidades en el empleo, en el acceso a bienes culturales, en la forma en que se relacionan con el medio y con otros grupos sociales, el Sistema Educativo Nacional debe encontrar formas más efectivas para que los estudiantes adquieran más y mejores aprendizajes.

En Español

- No se encuentran diferencias significativas en la población total entre 2005 y 2008.
- La única diferencia significativa encontrada al desagregar la población por modalidades se encuentra en Telesecundaria, en que hubo un aumento de 8 puntos (la doceava parte de una desviación estándar). Esta diferencia de resultados le significa a la Telesecundaria que en 2008 su puntuación promedio no se encuentre en el nivel Por debajo del básico, como en 2005, sino en el nivel Básico. A esta diferencia parecen contribuir sólo las mujeres, pues al desagregar por sexo, las mujeres de Telesecundaria muestran un aumento significativo de 12 puntos, mientras que los hombres muestran un aumento no significativo de dos puntos.
- A pesar de esta mejora en Telesecundaria, la mitad de sus estudiantes se encuentran aún en el nivel Por debajo del Básico. A nivel nacional, poco más de la tercera parte de los estudiantes de tercero de secundaria continúan en el nivel Por debajo del básico en Español.



- La brecha entre promedios de la modalidad más desfavorecida, Telesecundaria, y la que tiene mayor puntuación, Privada, disminuyó de 2005 a 2008 en un quinto de desviación estándar. A ello contribuyeron tanto el aumento significativo en el promedio de Telesecundaria como una disminución de 11 puntos en la media estimada de la modalidad Privada, aunque ésta no es estadísticamente significativa.
- También disminuyó la brecha entre la media de Telesecundaria y la media nacional, en un décimo de desviación estándar. Además, la Telesecundaria es la única modalidad en la que disminuyó (ligera) el coeficiente de variación en el logro en Español, es decir, disminuyeron las diferencias en el aprendizaje de esta asignatura entre quienes se encuentran en dicho grupo.
- Al analizar a la población por sexo y modalidad simultáneamente, las únicas diferencias significativas fueron disminución tanto a nivel nacional como en modalidad privada de los hombres, de siete y 25 puntos, respectivamente. Quizá la primera es causada en buena parte por la segunda, puesto que ésta es de magnitud considerable (una cuarta parte de la desviación estándar de la población).
- Como consecuencia de las diferencias en el logro en Español por sexo en el periodo, la distancia en las puntuaciones promedio entre mujeres y hombres, que se ha observado siempre a favor de las primeras en todas las aplicaciones Excale de Español, aumentó nueve puntos de 2005 a 2008. Esta distancia no preocupa, ya que, además de ser pequeña, no existen evidencias de que de manera generalizada en nuestro país existan prácticas sexistas que excluyan a los hombres del aprendizaje del español, además de que parece haber una tendencia en los resultados de estudios internacionales de aprendizaje en esta área a que las mujeres tengan mejor desempeño (por ejemplo, Van de Gaer *et al.*, 2009; Bornstein *et al.* 2004; Karrass *et al.*, 2002) y éste no parece estar ligado a discriminación por género.
- La gran brecha entre edad y extraedad (de alrededor de 60 puntos) se conservó sin cambios de 2005 a 2008. Es un pendiente de la educación disminuir esta brecha, buscando alternativas a la repetición de cursos que logren compensar de manera efectiva el aprendizaje de los estudiantes en situación de extraedad.
- Al desagregar los grupos de edad por modalidad, encontramos las siguientes diferencias significativas entre 2005 y 2008: en la modalidad General, los estudiantes con extraedad obtienen 13 puntos menos en promedio en Español en 2008; en Telesecundaria, los estudiantes en edad normativa aumentan ocho puntos en 2008; y en la modalidad Privada, los estudiantes en edad normativa disminuyen 11 puntos en el mismo año. Estos resultados apuntan hacia un mejor apoyo a los



estudiantes en situación de extraedad en Telesecundaria en 2008 que en 2005?, ¿apuntan hacia un manejo menos pertinente de la situación de estos estudiantes en 2008 que en 2005 en las modalidades General y Privada? Estas preguntas requieren un trabajo de indagación focalizado para responderse adecuadamente.

En Matemáticas

- No se encuentran diferencias significativas entre 2005 y 2008 en el promedio de logro a nivel nacional ni al desagregar la población por modalidades. Por otro lado, aunque no son cambios estadísticamente significativos, las medias de todas las modalidades son mayores en 2008 que en 2005. Esto, en el caso de la modalidad General, implica que su promedio abandone el nivel Por debajo del básico y se ubique en el nivel Básico.
- La distribución de los alumnos a nivel nacional y por modalidades en los cuatro niveles de logro es muy similar en los dos años de aplicación; en la distribución nacional más de la mitad de los estudiantes se encuentran en el nivel Por debajo del básico, y en telesecundarias seis de cada 10. Al igual que en Español, es necesario que el Sistema Educativo Nacional encuentre formas efectivas de mejorar los aprendizajes de los estudiantes, pues los cambios inerciales (probablemente debidos a cambios en las condiciones sociales de la población) o los que están localizados en algunas modalidades no son suficientes para mejorar la calidad del sistema educativo en su conjunto.
- A diferencia de Español, donde la modalidad Telesecundaria disminuyó brechas, en Matemáticas la brecha entre la modalidad más desfavorecida, Telesecundaria, y la que tiene mayor puntuación, Privada, es prácticamente la misma en los dos años del estudio (alrededor de una desviación estándar). La brecha entre Telesecundaria y la media nacional también se encuentra sin diferencias en ambos años (es de alrededor de un cuarto de desviación estándar).
- La diferencia de puntuaciones entre mujeres y hombres, pequeña pero estadísticamente significativa, es muy similar en ambos años del estudio.
- La brecha entre estudiantes en edad normativa y en situación de extraedad se conservó en 54 puntos, sin variaciones entre los dos años de estudio. Esta brecha es grande (poco más de media desviación estándar), por lo que, al igual que en Español, es importante hacer un análisis minucioso sobre la conveniencia de conservar el esquema de repetición de grados o considerar nuevas alternativas que lleven a que los estudiantes rezagados en conocimientos logren alcanzar resultados comparables a los de los demás alumnos al finalizar la secundaria.

- Al desagregar simultáneamente a la población por edad y modalidad, se encuentran diferencias significativas entre 2005 y 2008 en los estudiantes en edad normativa de Telesecundaria que mejoraron en este periodo de tiempo.

Es interesante que la diferencia más consistente en Español, una pequeña mejora en Telesecundaria, es también la única diferencia significativa entre años al analizar los resultados de Matemáticas, aunque en este segundo caso es sólo en los estudiantes en edad normativa. A partir de estos análisis, es posible considerar que en las telesecundarias hubo cambios positivos durante el periodo; otra posible explicación, que podría ser o no complementaria, sería la existencia de cambios en los contextos de los alumnos. A partir de los cuestionarios de opinión aplicados a los alumnos junto con los Excale, se observa que el nivel de escolaridad de la madre que cambió más en las modalidades General y Técnica (el nivel Bachillerato), aumentó alrededor de tres puntos, mientras que el nivel de escolaridad de la madre que cambió más en telesecundarias (el nivel Secundaria) aumentó siete puntos. En los tres casos, el aumento implicó la disminución de casos en los niveles inferiores. Los niveles de escolaridad de las madres de los estudiantes de modalidad Privada no cambiaron en este intervalo de tiempo. Así pues, los niveles de escolaridad de las madres de los estudiantes de telesecundarias aumentaron en mayor proporción que en otras modalidades educativas en el periodo. La escolaridad del padre se movió de manera similar, también en mayor medida en Telesecundaria y sin cambio en la modalidad Privada. Estos datos pueden hacer pensar que el aumento pequeño pero constante en la escolaridad de la población, y específicamente de los padres de los estudiantes de secundaria, como indicador de mejores condiciones de vida y de un ambiente familiar con mayor capital cultural escolar, contribuye a los pequeños cambios positivos que se encontraron en este informe comparativo.

Sobre la información por entidad

Como se menciona en el apartado sobre desagregación por entidades, la información presentada no debe utilizarse para compararlas entre sí; se carece de elementos de contexto sobre las entidades, que serían necesarios para poder



hacer juicios de valor sobre el trabajo educativo en las mismas. Baker *et al.* (2010) señalan que la educación es un trabajo acumulativo y complejo en el que hay múltiples influencias sobre el aprendizaje de los alumnos. Por ello, el uso de los datos aquí se limita a mostrar cruces de modalidad-asignatura que pueden representar mayores fortalezas o debilidades. Estos datos deben ser puestos en contexto, junto con otros disponibles sobre una entidad dada, de modo que se lleven a cabo análisis que ofrezcan diagnósticos certeros por estado.

Para terminar

Este trabajo, que muestra datos de dos años no consecutivos, aún no nos permite hablar de tendencias. La próxima aplicación del Excale de tercero de secundaria, que se hará a mediados de 2012, dará un tercer punto de comparación que enriquecerá los análisis. El resultado de la comparación de los tres puntos en el tiempo se reportará a mediados de 2013.

Bibliografía

- Backhoff, E. y Díaz, M.A. (2005). *Plan General de Evaluación del Aprendizaje*. México: Instituto Nacional para la Evaluación de la Educación.
- Backhoff, E., Andrade, E., Sánchez, A., Peon, M. y Bouzas, A. (2006). *El aprendizaje del Español y las Matemáticas en la educación básica de México*. México: Instituto Nacional para la Evaluación de la Educación.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R.L. Ravitch, D., Rothstein, R., Shavelson, R.J. y Shepard, L. A. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers*. Washington DC: Economic Policy Institute.
- Beaton, A. E., y Zwick, R. (1990). The Effect of Changes in the National Assessment: Disentangling the NAEP 1985–86 Reading Anomaly (No. 17-TR-21). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.
- Bornstein, M. H., Hahn, C. y Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, Octubre 2004 24: 267-304.
- Hambleton, R.K., Gonzalez, E., Plake, B. y Ponocny, I. (2005). *Technical Review of PISA*. Published in OECD (2006), Call for Tender for an international contractor for the implementation of the PISA 2009.
- INEE (2006). *Plan maestro de desarrollo 2007-2014*, México: INEE.
- Joint Committee on Testing Practices (2004). *Code of Fair Testing Practices in Education*. Washington DC: Joint Committee on Testing Practices.
- Jornet, J. y Backhoff, E. (2006). Manual técnico. *Establecimiento de niveles de competencia*. México: INEE.
- Karrass, J., Braungart-Rieker, J. M., Mullins, J. y Lefever, J. (2002). Processes in Language Acquisition: The Roles of Gender, Attention, and Maternal Encouragement of Attention over Time. *Journal of Child Language*, 29(3), 519-543.
- Mislevy, R. J. (1992). *Linking Educational Assessments: Concepts, issues, methods, and prospects*. ETS.
- Organisation for Economic Co-operation and Development (2009). *PISA Data Analysis Manual*. SPSS, Second Edition. OECD.
- Qian, J. (2006). *Linking 2005 NAEP Science Assessment through Bridge Samples*. Trabajo presentado en la IOMW Meeting. ETS.
- Sánchez, A. y Andrade, E. (coords.) (2009). El aprendizaje en tercero de secundaria en México. *Informe sobre los resultados de Excale 09, aplicación 2008. Español, Matemáticas, Biología y Formación cívica y ética*. México: Instituto Nacional para la Evaluación de la Educación.
- SEP (1993). Plan y programas de estudios 1993. Educación Básica. Primaria. México: SEP.
- Van de Gaer, E., Pustjens, H., Van Damme, J. y De Munter, A. (2009). School Engagement and Language Achievement: A Longitudinal Study of Gender Differences across Secondary School. Merrill-Palmer Quarterly: *Journal of Developmental Psychology*, 55(4), 373-405.

Índice de tablas

11 Tabla I.

Número de escuelas y matrícula en tercero de secundaria a nivel nacional y por tipo de servicio y sostenimiento en los ciclos 2004-2005 y 2007-2008.

11 Tabla II.

Número de escuelas y alumnos en las muestras utilizadas para el estudio comparativo 2005-2008.

13 Tabla III.

Niveles de logro genéricos.

18 Tabla IV.

Bloques de reactivos para la aplicación 2005 del Excale 09.

19 Tabla V.

Cambios en la cantidad de reactivos Excale 09 entre 2005 y 2008, por asignatura.

20 Tabla VI.

Ejemplo de diseño matricial propuesto para Excale 09 aplicación 2008.

23 Tabla VII.

Cuantiles de la prueba puente y muestra operacional Excale 09 aplicación 2008.

25 Tabla VIII.

Estructura del Excale 09: Español.

26 Tabla IX.

Definición de los niveles de logro en Español.

28 Tabla X.

Puntuaciones medias por modalidad educativa en Español.

31 Tabla XI.

Dispersión de puntajes en Español, por modalidad.

33 Tabla XII.

Puntuaciones medias en Español por modalidad y sexo.

35 Tabla XIII.

Puntuaciones medias en Español por modalidad educativa y edad.

37 Tabla XIV.

Estructura del Excale 09 de Matemáticas.

38 Tabla XV.

Niveles de logro: Matemáticas.

41 Tabla XVI.

Puntuaciones medias en Matemáticas por modalidad educativa.

43 Tabla XVII.

Dispersión de puntajes en Matemáticas, por modalidad.

45 Tabla XVIII.

Puntuaciones medias en Matemáticas por modalidad y sexo.

47 Tabla XIX.

Puntuaciones medias en Matemáticas por edad y modalidad.

49 Tabla XX.

Entidades por arriba de la media nacional por asignatura y por modalidad educativa, 2005 y 2008.

49 Tabla XXI.

Entidades por debajo de la media nacional por asignatura, por modalidad educativa, 2005 y 2008.

Índice de gráficas

- | | |
|---|--|
| <p>9 Figura 1.
Plan de Evaluación del logro con Excale 2004-2016.</p> <p>21 Figura 2.
Esquema para la aplicación y comparación de pruebas Excale 2005-2008.</p> <p>27 Figura 3.
Mapa parcial de reactivos: Español.</p> <p>29 Figura 4.
Comparación de las puntuaciones medias por modalidad educativa en Español.</p> <p>30 Figura 5.
Porcentaje de estudiantes por nivel de logro educativo y modalidad educativa en Español.</p> <p>32 Figura 6.
Comparación de las puntuaciones medias en Español por sexo.</p> <p>33 Figura 7.
Porcentaje de estudiantes por nivel de logro educativo en Español y sexo.</p> <p>34 Figura 8.
Comparación de las puntuaciones medias en Español por edad.</p> | <p>40 Figura 9.
Mapa parcial de reactivos: Matemáticas¹.</p> <p>42 Figura 10.
Comparación de las puntuaciones medias en Matemáticas por modalidad educativa.</p> <p>43 Figura 11.
Porcentaje de estudiantes por nivel de logro educativo en Matemáticas por modalidad educativa.</p> <p>44 Figura 12.
Comparación de las puntuaciones medias en Matemáticas por sexo.</p> <p>45 Figura 13.
Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y sexo.</p> <p>46 Figura 14.
Comparación de las puntuaciones medias en Matemáticas por edad.</p> <p>46 Figura 15.
Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y edad.</p> |
|---|--|

¹ Si el lector desea conocer la información de la totalidad de los reactivos que conformaron esta prueba puede consultar el Anexo 3 de Sánchez y Andrade (2009).

Anexos

Anexo 1. Tablas complementarias de Español

Es posible utilizar los datos de la tabla A para localizar diferencias significativas en los porcentajes de logro entre grados, debido a que las muestras son independientes.

Tabla
A

Porcentaje de estudiantes por nivel de logro educativo y modalidad educativa en Español

Modalidad educativa	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Nacional	32.7	(0.6)	35.9	(0.8)	38.3	(0.7)	35.9	(0.7)	23.7	(0.5)	22.1	(0.6)	5.3	(0.2)	6.1	(0.5)
General	29.7	(1.1)	34.7	(1.8)	40.7	(1.1)	36.8	(1.3)	25.0	(0.9)	22.8	(1.2)	4.6	(0.4)	5.7	(0.8)
Técnica	31.1	(0.9)	33.7	(0.9)	39.8	(0.8)	36.8	(0.9)	24.7	(0.8)	23.4	(1.0)	4.5	(0.3)	6.0	(0.6)
Telesecundaria	51.1	(1.1)	50.1	(1.5)	35.6	(1.1)	35.5	(1.8)	12.1	(0.7)	13.0	(0.9)	1.2*	(0.3)	1.4	(0.3)
Privada	8.1	(0.5)	11.9	(1.3)	27.4	(1.0)	27.9	(1.7)	42.4	(0.7)	39.0	(1.7)	22.2	(1.1)	21.2	(1.4)

* Estimación cuyo coeficiente de variación excede a 20%.

Tabla
B

Puntuaciones medias en Español por sexo

Sexo	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)
Masculino	488	(1.4)	482	(2.6)	-7	(3.0)
Femenino	511	(1.4)	514	(2.1)	2	(2.6)
Diferencia*	-23	(1.4)	-32	(2.5)	-9	(2.9)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

Tabla
C

Porcentaje de estudiantes por nivel de logro educativo en Español y sexo

Sexo	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Hombres	37.2	(0.7)	42.6	(1.2)	37.7	(0.8)	34.4	(0.8)	20.7	(0.6)	18.2	(0.8)	4.4	(0.2)	4.7	(0.6)
Mujeres	28.2	(0.6)	29.5	(1.0)	38.9	(0.9)	37.2	(1.1)	26.6	(0.6)	25.9	(1.0)	6.3	(0.3)	7.4	(0.6)

Tabla
D

Puntuaciones medias en Español por edad

Sexo	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)
Edad normativa	509	(1.4)	507	(2.1)	-1	(2.5)
Extraedad	449	(2.3)	450	(3.0)	0	(3.5)
Diferencia*	60	(2.6)	58	(2.8)	2	(3.9)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

Tabla
E

Porcentaje de estudiantes por nivel de logro educativo en Español y edad

Alumnos	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Edad normativa	29.2	(0.6)	32.1	(0.9)	39.1	(0.7)	36.7	(0.8)	25.8	(0.5)	24.5	(0.7)	5.9	(0.3)	6.7	(0.5)
Extraedad	53.2	(1.1)	56.1	(1.5)	33.6	(1.3)	32.0	(1.8)	11.4	(0.8)	9.2	(1.0)	1.7	(0.3)	2.7	(0.5)

Anexo 2. Tablas complementarias de Matemáticas

Es posible utilizar los datos de la tabla F para localizar diferencias significativas en los porcentajes por nivel entre años, debido a que las muestras son independientes.

Tabla
F

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y modalidad educativa

Alumnos	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Nacional	51.1	(0.6)	51.9	(1.0)	29.5	(0.5)	27.5	(0.7)	18.0	(0.4)	18.5	(0.8)	1.4	(0.1)	2.1	(0.3)
General	50.5	(1.3)	50.5	(1.8)	30.5	(1.0)	29.0	(1.4)	17.9	(0.8)	18.9	(1.4)	1.1	(0.2)	1.6*	(0.4)
Técnica	52.0	(0.8)	54.0	(1.2)	30.2	(0.5)	27.5	(0.8)	16.9	(0.6)	16.8	(0.9)	0.9	(0.2)	1.7*	(0.4)
Telesecundaria	62.1	(1.1)	62.1	(1.4)	26.0	(0.8)	23.4	(1.4)	11.4	(0.6)	13.5	(1.3)	0.5*	(0.2)	0.9*	(0.4)
Privada	23.7	(0.9)	24.5	(1.7)	31.0	(0.9)	29.3	(1.5)	38.4	(0.9)	36.5	(1.8)	7.0	(0.6)	9.7	(1.4)

* Estimación cuyo coeficiente de variación excede a 20%.

Tabla
G

Puntuaciones medias en Matemáticas por sexo

Sexo	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)
Masculino	504	(1.5)	508	(2.6)	5	(2.9)
Femenino	498	(1.7)	501	(2.4)	3	(2.8)
Diferencia*	6	(1.4)	8	(2.7)	-1.73	(3.0)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

Tabla
H

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y sexo

Sexo	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Hombres	49.7	(0.6)	50.6	(1.2)	29.9	(0.6)	27.5	(1.2)	18.8	(0.5)	19.2	(1.1)	1.6	(0.2)	2.7	(0.5)
Mujeres	51.7	(0.8)	53.1	(1.3)	29.4	(0.6)	27.4	(1.2)	17.6	(0.6)	18.0	(1.0)	1.3	(0.1)	1.5	(0.3)

Tabla
I

Puntuaciones medias en Matemáticas por edad

Alumnos	2005		2008		Diferencia*	
	Media	(EE)	Media	(EE)	Media	(EE)
Edad normativa	508	(1.5)	513	(2.4)	5	(2.6)
Extraedad	454	(1.9)	459	(2.7)	6	(3.3)
Diferencia*	54	(2.2)	54	(3.8)	1	(4.4)

* Las diferencias pueden no ser exactas debido al redondeo.
Las diferencias significativas están marcadas con negritas.

Tabla
J

Porcentaje de estudiantes por nivel de logro educativo en Matemáticas y edad

Alumnos	Por debajo del básico				Básico				Medio				Avanzado			
	2005		2008		2005		2008		2005		2008		2005		2008	
	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)	%	(EE)
Edad normativa	47.7	(0.6)	48.1	(1.1)	30.9	(0.5)	29.1	(0.8)	19.8	(0.5)	20.5	(0.9)	1.6	(0.1)	2.4	(0.3)
Extraedad	70.3	(1.0)	71.1	(1.6)	22.0	(0.9)	19.7	(1.8)	7.4	(0.5)	8.4	(0.9)	0.3	(0.1)	0.8	(0.3)



Resultados de
evaluaciones

ESTUDIO COMPARATIVO DEL APRENDIZAJE EN
TERCERO DE SECUNDARIA EN MÉXICO
2005-2008. ESPAÑOL Y MATEMÁTICAS
Se terminó de imprimir en septiembre de 2012
en los talleres de IEPSA. En su formación se utilizaron
las familias tipográficas: Helvetica Neue LT Std, The Serif.
Tiraje: 1 000 ejemplares.



Obtenga una copia digital de esta publicación, sin costo.



Visite nuestro portal:
<http://www.inee.edu.mx>



Comuníquese con nosotros:
informes@inee.edu.mx

Las pruebas Excale (Exámenes de la Calidad y el Logro Educativos), diseñadas y aplicadas por el Instituto Nacional para la Evaluación de la Educación, miden el logro escolar de los estudiantes de educación básica en su conjunto, con base en el currículo que proponen los planes y programas de estudio en México; su diseño permite medir el dominio de los estudiantes en una disciplina en particular y compararlo a lo largo del tiempo.

El informe que usted tiene en sus manos contrasta los resultados de los Excale de tercero de secundaria en dos momentos de aplicación (2005 y 2008), en las áreas de Español y Matemáticas. De esta manera, se continúa la labor iniciada en 2005 con el *Estudio Comparativo de la Educación Básica en México: 2000-2005*, que entonces utilizó las Pruebas de Estándares Nacionales de la Dirección General de Evaluación de la SEP. Posteriormente, el INEE publicó el *Estudio Comparativo del Aprendizaje en Sexto de Primaria en México 2005-2007*, que se basó en las pruebas Excale.

Sin ahondar en términos técnicos, con el propósito de que este texto sea accesible a un público extenso, se incluye el método utilizado para equiparar las pruebas Excale de diferentes años y del mismo grado escolar.

El lector encontrará en este volumen diferentes elementos para la reflexión, surgidos de la desagregación de los datos por modalidad, sexo y edad.

Con los informes de resultados de los Excale —y otros más como la prueba PISA aplicada en los países miembro de la OCDE— el INEE espera contribuir a la discusión y construcción de propuestas que mejoren los aprendizajes de los estudiantes mexicanos.



Resultados de
evaluaciones

Logro escolar

Aplicaciones de la prueba Excale