

Comparación de los resultados de las Pruebas Nacionales de educación básica: 2003 y 2004

EDUARDO BACKHOFF ESCUDERO *



En este trabajo se analizan los alcances y limitaciones de la comparación de resultados de las Pruebas Nacionales que aplicó el Instituto Nacional para la Evaluación de la Educación en junio de 2003 y 2004, con el propósito de conocer las habilidades lectoras y matemáticas de los estudiantes de sexto grado de primaria y tercero de secundaria. Los resultados de las pruebas se comparan y se analizan por estrato y modalidad educativos, y por entidad federativa. Los datos muestran, con relativa estabilidad, que las escuelas privadas son las que obtienen mejores puntuaciones, mientras que las escuelas indígenas y telesecundarias obtienen las puntuaciones más bajas. Sin embargo, cuando se comparan las puntuaciones de las entidades federativas en el tiempo, se observa que los resultados se comportan con mayor inestabilidad. Se concluye que no es posible conocer el avance que tuvo el país a lo largo del tiempo, debido a las diferencias de los instrumentos utilizados.

In this article the author analyzes the achievements and limitations shown by the comparison of results of the National Tests implemented by the National Institute for Educational Assessment in June of 2003 and 2004 in order to have a grip on the reading and mathematics abilities of pupils who attend the sixth grade of basic education and the third grade of secondary education. The results of the tests are compared and analyzed by stratum, educational modalities and federal entity. The data show, with a relative stability, that the private institutions get the best score, whereas the indigenous schools and the distance secondary institutions obtain the lowest marks. However, when the point of comparison takes into account the federal entity in a diachronic perspective, the results are less stable, which leads the conclusion that it is not possible to know the actual progress experienced by the country along the years because of the significant differences in the instruments that have been used.

Educación básica / Pruebas nacionales / Evaluación / Habilidad lectora /
Habilidad matemática / Análisis cuantitativo

Basic education / National tests / Assessment / Reading ability / Mathematics ability / Quantitative analysis

Recibido: 25.02.2005 / aprobado: 6.06.2005

Uno de los propósitos del Instituto Nacional para la Evaluación de la Educación (INEE) es proveer información relevante sobre los resultados de aprendizaje de los estudiantes de educación básica, que ayude a mejorar la calidad de los servicios educativos que se ofrecen en el país (Poder Ejecutivo, 2002). Dicho propósito se ha venido cumpliendo por medio de las publicaciones de los informes nacionales y la participación de México en los estudios internacionales (INEE, 2003 y 2004).

La información de las evaluaciones nacionales e internacionales que han comenzado a difundirse en los últimos años ha despertado gran interés entre autoridades y público en general. Además de innegables factores publicitarios, buenas razones explican ese interés: la educación es un factor indispensable del desarrollo nacional y del bienestar individual de las personas.

La afirmación anterior toma importancia cuando consideramos las dimensiones del sistema educativo de un país, como es el caso de México, que tiene más de 31 millones de alumnos, de los cuales 24 millones asisten a algunos de los grados de la educación básica, en más de 205 mil escuelas distribuidas a lo largo y ancho del país (INEE, 2004).

Así, no sorprende que al público mexicano le importe saber si la educación que ofrece su sistema educativo es de mejor o peor calidad, cuando éste se compara consigo mismo a lo largo del tiempo. La competitividad de México se verá afectada, favorable o desfavorablemente, por la calidad de la educación que reciban los alumnos, futuros ciudadanos y productores de bienes y servicios (Backhoff, 2004).

Sin embargo, no es fácil comparar de manera precisa la calidad en el tiempo de un sistema educativo; en ausencia de instrumentos y metodologías apropiadas, tales comparaciones son imposibles de realizar. Afortunadamente, hace ya casi medio siglo comenzaron a desarrollarse proyectos de evaluación internacionales cuyos paulatinos aportes permiten hacer comparaciones consistentes, válidas y confiables, aplicando pruebas de rendimiento equivalentes a muestras representativas de estudiantes (Backhoff y Martínez-Rizo, 2004).

La experiencia de México en este rubro no ha sido del todo satisfactoria. Aunque se han querido hacer comparaciones longitudinales, atendiendo al reclamo de académicos y de los medios de comunicación, los instrumentos y procedimientos nacionales utilizados hasta el momento no han permitido que el INEE proporcione dicha información (Martínez-Rizo, 2004).

Por lo anterior, el propósito de este trabajo es doble: por un lado, analizar y discutir las limitaciones de la comparabilidad de los resultados obtenidos con las Pruebas Nacionales de 2003 y 2004, por el otro, comparar los resultados de las evaluaciones efectuadas en 2003 y 2004 en dos niveles de agregación: 1) por estrato y modalidad educativos y 2) por entidad federativa.

* Doctor en Educación por la Universidad Autónoma de Aguascalientes, maestro en Educación por la Universidad de Washington. Fue director del Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California (UABC). Es miembro fundador del Consejo Asesor Externo del Centro Nacional para la Evaluación de la Educación Superior (CENEVAL) y consejero técnico fundador del mismo Centro. Actualmente es investigador del Instituto de Investigación y Desarrollo Educativo de la UABC. Correo electrónico: eduardo.backhoff@ineemexico.org

ALCANCES Y LIMITACIONES EN LA COMPARACIÓN DE RESULTADOS DE LAS PRUEBAS NACIONALES

Martínez-Rizo (2004) ha explicado por qué las Pruebas de Estándares Nacionales que la Dirección General de Evaluación (DGE) de la Secretaría de Educación Pública (SEP) construyó de 1998 a 2003 y la Prueba Nacional que el INEE elaboró en 2004 (con el mismo esquema de las pruebas de la DGE) no cumplen con los elementos técnicos para su comparación. Sin embargo, no se han hecho las suficientes precisiones como para aclarar el punto a satisfacción. Veamos los argumentos de mayor peso para el caso de las pruebas de 2003 y 2004.

Por un lado, es importante conocer las características de las pruebas utilizadas que permiten un cierto grado de comparación en ambas evaluaciones:

- *Sus contenidos son semejantes.* En ambos casos se evalúan los mismos contenidos y grados escolares: comprensión lectora y matemáticas de sexto grado de primaria y tercero de secundaria. Asimismo, los reactivos están diseñados para evaluar habilidades lectoras y matemáticas, más que conocimientos curriculares.
- *No hubo cambios de referentes curriculares en los tiempos evaluados.* El Sistema Educativo Nacional no implementó cambios en los planes y programas de estudios de 2003 a 2004 en las habilidades evaluadas.
- *Los procedimientos de muestreo fueron los mismos.* El marco y el diseño muestral utilizado en ambas evaluaciones fueron los de 1998. Por lo tanto, las poblaciones estudiadas fueron equivalentes en tamaño y distribución.

- *Las evaluaciones se realizaron en el mismo mes del año.* En ambos casos las aplicaciones ocurrieron en junio del ciclo escolar correspondiente.
- *Se utilizaron los mismos procedimientos de aplicación y lectura de resultados.* En todos los casos se utilizaron los mismos protocolos de aplicación de pruebas y los mismos procedimientos de lectura óptica de las hojas de respuesta.
- *Las pruebas contienen un número suficiente de reactivos ancla.*¹ En teoría, para que dos pruebas puedan ser equiparables deben contener al menos 20% de reactivos comunes, que fue el caso para las cuatro pruebas aplicadas en 2003 y 2004 (véase cuadro 1).
- *Las técnicas de calibración y escalamiento fueron las mismas.* En todos los casos se calibraron las pruebas con el modelo de Rasch, se descartaron los reactivos que no cumplieron con los criterios establecidos y se escalaron las calificaciones en un rango de 200 a 800 puntos, con una media de 500 y una desviación estándar de 100.

Por otro lado, las diferencias y deficiencias que dificultan y no permiten las comparaciones de los resultados de las pruebas son:

- *Las especificaciones de los contenidos no se encuentran bien definidas.* Ambas pruebas carecen de un marco teórico o conceptual robusto donde, con toda claridad, se explicita su diseño, estructura y composición temática.
- *No se diseñaron con una metodología ex profeso para su equiparación.*² Por lo anterior, las pruebas no tuvieron la misma extensión (número de preguntas), la misma dificultad (proporción

- de respuestas correctas), ni confiabilidades³ similares (véase cuadro 1).
- *Los reactivos ancla de las pruebas no tuvieron la misma ubicación.* Para una equiparación correcta es necesario que los reactivos ancla se ubiquen en la misma posición en las pruebas que se desea equiparar, condición que no se cuidó para el caso de estas evaluaciones.
 - *Ninguna prueba se pilotó antes de su aplicación.* Al no probarse los reactivos con anticipación, es imposible seleccionarlos de tal forma que sus características psicométricas y, por lo tanto, sus resultados, sean equiparables.

- *Se utilizaron las mismas escuelas seleccionadas desde 1998.* Aunque esta condición pudiera parecer una ventaja para la comparación, en realidad es una debilidad metodológica, ya que las escuelas se “acostumbran” a ser evaluadas año con año, lo cual puede viciar los resultados de la evaluación.
- La DGE estuvo a cargo de la evaluación de 2003, mientras que el INEE coordinó la de 2004. Al producirse este cambio en la administración de las Pruebas Nacionales, también pudo ocurrir un cambio positivo o negativo en la actitud e interés de las escuelas evaluadas.

CUADRO 1 • Características psicométricas de las Pruebas Nacionales de 2003 y 2004.

Propiedades	6° de primaria				3° de secundaria			
	Comprensión lectora		Matemáticas		Comprensión lectora		Matemáticas	
	2003	2004	2003	2004	2003	2004	2003	2004
Extensión	58	42	46	44	62	45	48	44
Dificultad*	0.24	0.20	0.13	0.21	0.37	0.24	0.18	0.21
Confiabilidad**	0.85	0.86	0.78	0.84	0.89	0.86	0.68	0.83
Reactivos ancla	16		10		12		11	

* Calculada con la fórmula Alfa de Cronbach

** Proporción de estudiantes que respondieron correctamente los reactivos.

Por lo anterior, técnicamente hablando, los resultados de las Pruebas Nacionales de 2003 y 2004 no son equiparables, razón por la cual no podemos observar si hubo un progreso o retroceso en este lapso. Sin embargo, sí es posible comparar los resultados más generales y analizar las conclusiones a las que se llega en cada una de las evaluaciones.

COMPARACIÓN DE LOS RESULTADOS DE LAS PRUEBAS NACIONALES 2003 Y 2004

Como ya se mencionó, en ambas evaluaciones el diseño de las muestras se obtuvo con el mismo marco muestral. El cuadro 2 muestra los distintos estratos y modalidades educativos que estuvieron

representados en el ámbito nacional en los estudios de 2003 y 2004. Como se podrá apreciar, las poblaciones estudiadas

fueron equivalentes en tamaño y distribución, con diferencias menores que no son significativas estadísticamente.

CUADRO 2 • Muestras usadas en las pruebas 2003 y 2004.

Estratos y modalidades	Escuelas		Alumnos	
	2003	2004	2003	2004
6° de primaria	3 119	3 150	48 087	53 129
Privadas (urbanas)	143	140	3 248	3 562
Públicas	2 976	3 010	44 839	49 567
Urbanas	901	906	22 319	25 286
Rurales	1 095	1 100	14 033	14 900
Cursos comunitarios	289	292	550	716
Escuelas indígenas	694	712	7 937	8 665
3° de secundaria	1 175	1 184	38 760	37 973
Privadas	74	76	2 276	2 230
Públicas	1 101	1 108	36 484	35 743
Generales	559	564	19 062	18 734
Técnicas	480	482	16 298	15 891
Telesecundarias	62	62	1 124	1 118

Fuente: INEE (2003, 2004), La calidad de la educación básica en México.

A continuación se analizan los resultados obtenidos en ambas evaluaciones. En primer lugar compararemos las diferencias entre estratos y modalidades educativos, para después comparar los resultados de las 32 entidades federativas del país.

COMPARACIÓN ENTRE ESTRATOS Y MODALIDADES EDUCATIVOS

Los cuadros 3 y 4 presentan los resultados promedio de los estudiantes por estrato y modalidad educativos para los grados sexto de primaria y tercero de secundaria, respectivamente.

Es importante señalar que en ambos cuadros los resultados de 2004 son mejores que los del 2003 en todos los casos; condición que nos llevaría a pensar que hubo un avance sustancial en el país en el lapso de un año, tanto en primaria como en secundaria. Sin embargo, cabe recordar que como las pruebas de ambas aplicaciones no se pueden igualar y utilizan sus propias escalas, esta comparación y deducción son inadecuadas, ya que los resultados en términos de sus puntuaciones no son equivalentes.

No obstante, sí podemos comparar los resultados en cada aplicación para después

CUADRO 3 • Medias y errores estándar de las puntuaciones de alumnos de sexto grado de primaria en comprensión lectora y matemáticas, por estrato, 2003 y 2004.

Estrato	Comprensión lectora				Matemáticas			
	2003		2004		2003		2004	
	Media	Error estándar	Media	Error estándar	Media	Error estándar	Media	Error estándar
Privadas	547.48	3.69	567.99	3.67	459.32	4.16	474.68	3.32
Urbanas públicas	475.31	1.78	497.33	1.57	405.69	1.76	425.02	1.45
Rurales públicas	439.26	2.01	462.33	1.79	386.44	1.74	399.60	1.73
Cursos comunitarios	425.66	4.56	450.59	3.61	377.93	3.80	385.79	3.82
Educación indígena	405.08	3.17	424.85	2.67	372.48	2.72	360.07	2.67
Todas las escuelas	466.30	1.24	488.49	1.10	402.00	1.20	417.98	1.03

Fuente: INEE (2003, 2004), La calidad de la educación básica en México.

CUADRO 4 • Medias y errores estándar de las puntuaciones de alumnos de tercer grado de secundaria en comprensión lectora y matemáticas, por modalidad, 2003 y 2004.

Modalidad educativa	Comprensión lectora				Matemáticas			
	2003		2004		2003		2004	
	Media	Error estándar	Media	Error estándar	Media	Error estándar	Media	Error estándar
Privadas	547.48	3.69	567.99	3.67	459.32	4.16	474.68	3.32
Urbanas públicas	475.31	1.78	497.33	1.57	405.69	1.76	425.02	1.45
Generales	439.26	2.01	462.33	1.79	386.44	1.74	399.60	1.73
Técnicas	425.66	4.56	450.59	3.61	377.93	3.80	385.79	3.82
Telesecundarias	405.08	3.17	424.85	2.67	372.48	2.72	360.07	2.67
Todas las escuelas	466.30	1.24	488.49	1.10	402.00	1.20	417.98	1.03

Fuente: INEE (2003, 2004), La calidad de la educación básica en México.

contrastarlos y observar sus similitudes y diferencias; en condiciones como la de las Pruebas de Estándares Nacionales esto es lo apropiado. Veamos con esta óptica lo que nos dice la información de ambos estudios mostrada en los cuadros 3 y 4.

En principio, se observa una consistencia de los resultados cuando comparamos las puntuaciones de los estratos y modalidades educativos. En ambas evaluaciones, tanto en comprensión lectora como en matemáticas, los estratos de las escuelas primarias que obtienen mejor puntuación en orden descendente son: privadas, urbanas públicas, rurales públicas, cursos comunitarios y educación indígena. Asimismo, en las escuelas secundarias también se observa una consistencia en los resultados

de las distintas modalidades educativas. En las dos aplicaciones se da el siguiente ordenamiento de las escuelas, de mayor a menor puntuación: privadas, generales, técnicas y telesecundarias.

Sin embargo, también es importante analizar el valor estadístico de estas diferencias. En el cuadro 5 se muestra la comparación de los resultados de los alumnos de sexto grado de primaria, por estrato educativo. El cuadro 6 muestra estas diferencias para los alumnos de tercer grado de secundaria, de acuerdo con la modalidad educativa de las escuelas. La palabra *sí* indica que hay diferencias significativas entre dos estratos o modalidades educativas (en las celdas donde convergen los renglones y las columnas) y la palabra *no*

CUADRO 5 • Diferencias significativas de los resultados de alumnos de sexto grado de primaria por estrato educativo, 2003 vs. 2004.

Modalidad educativa	Privadas		Urbanas públicas		Rurales públicas		Cursos comunitarios	
	2003	2004	2003	2004	2003	2004	2003	2004
Comprensión lectora								
Urbanas públicas	sí	sí	—	—	—	—	—	—
Rurales públicas	sí	sí	sí	sí	—	—	—	—
Cursos comunitarios	sí	sí	sí	sí	NO	NO	—	—
Educación indígena	sí	sí	sí	sí	sí	sí	NO	sí
Matemáticas								
Urbanas públicas	sí	sí	—	—	—	—	—	—
Rurales públicas	sí	sí	sí	sí	—	—	—	—
Cursos comunitarios	sí	sí	NO	sí	NO	NO	—	—
Educación indígena	sí	sí	sí	sí	sí	sí	NO	sí

Fuente: Base de datos del INEE; aplicaciones 2003 y 2004. Método: Bonferroni modificado.

Sí: Desempeño del estrato significativamente inferior respecto al comparado con el de la columna.

NO: Desempeño no significativamente diferente.

CUADRO 6 • Diferencias significativas en las medias de alumnos de tercer grado de secundaria por modalidad educativa, 2003 vs. 2004.

Modalidad educativa	Privadas		Generales		Técnicas	
	2003	2004	2003	2004	2003	2004
<i>Comprensión lectora</i>						
Generales	sí	sí	—	—	—	—
Técnicas	sí	sí	NO	NO	—	—
Telesecundarias	sí	sí	sí	sí	sí	sí
<i>Matemáticas</i>						
Generales	sí	sí	—	—	—	—
Técnicas	sí	sí	NO	NO	—	—
Telesecundarias	sí	sí	NO	NO	NO	sí

Fuente: Base de datos del INEE; aplicaciones 2003 y 2004. Método: Bonferroni modificado.

SÍ: Desempeño medio significativamente inferior respecto a la modalidad educativa de la columna.

NO: Desempeño medio no significativamente diferente.

—: No es necesaria la comparación.

indica que no hay diferencias significativas entre ellos. Las celdas sombreadas indican que en esos espacios es redundante (e innecesaria) la comparación.

En sexto grado de primaria, las diferencias en comprensión lectora en 2003 y 2004 son casi las mismas en prácticamente todos los estratos educativos, con excepción de las correspondientes a cursos comunitarios y educación indígena, donde se observa una discrepancia entre un año y otro. En el caso de matemáticas se observan dos discrepancias: 1) cursos comunitarios y escuelas urbanas públicas y 2) cursos comunitarios y educación indígena.

Por su parte, en tercer grado de secundaria se observa que en comprensión lectora no se presentó ninguna discrepancia entre 2003 y 2004, mientras que en matemáticas la única discrepancia se presen-

tó en las modalidades educativas de secundarias técnicas y telesecundarias.

Otra forma de conocer la estabilidad de los resultados es comparando las diferencias entre los estratos y modalidades con mayor y menor rendimiento en ambas aplicaciones (véanse cuadros 3 y 4). Para el caso de sexto grado de primaria, calculamos la diferencia de las medias de las escuelas privadas y las escuelas indígenas; para tercer grado de secundaria esta diferencia se calculó entre las secundarias privadas y las telesecundarias. El cuadro 7 muestra los resultados de esta comparación, donde se puede observar que son muy heterogéneas y asimétricas las diferencias en las puntuaciones de 2003 a 2004, tanto para las habilidades evaluadas como para los niveles escolares. Resulta claro que hubo mayores diferen-

cias en matemáticas que en comprensión lectora y en el nivel de secundaria que en el de primaria.

Es importante señalar que en todos los casos las diferencias entre las puntuaciones extremas fueron más grandes en 2004 que en 2003, lo que nos llevaría a pensar que las brechas educativas se han agrandado, o que hay brechas más grandes en secundaria que en primaria o entre una habilidad y otra. Los resultados de las Pruebas Nacionales no nos permi-

ten decir esto debido a que entre un año y otro, y entre un contenido y otro, las escalas son diferentes y no se pueden utilizar procedimientos de equiparación confiables.

Sin embargo, sí podemos decir que hay una consistencia en los ordenamientos de los estratos y modalidades educativos, y que las distancias entre ellos no se conservan equidistantes de un año a otro, debido a que se utilizaron instrumentos diferentes.

CUADRO 7 • Diferencia de medias entre los estratos y modalidades extremos en las evaluaciones de 2003 y 2004.

Grados (estratos y modalidades)	Pruebas	Diferencias en puntuaciones extremas		% de variación
		2003	2004	
6° de primaria (Escuelas privadas vs. indígenas)	Comprensión lectora	142	143	00.007
	Matemáticas	87	114	31.03
3° de secundaria (Secundarias privadas vs. telesecundarias)	Comprensión lectora	114	126	10.52
	Matemáticas	46	64	39.13

COMPARACIÓN ENTRE ENTIDADES FEDERATIVAS

Para comparar los resultados de las entidades federativas en 2003 y 2004, presentamos el cuadro 8, que muestra las diferencias significativas de las 32 entidades respecto a la media nacional, en las cuatro pruebas utilizadas. Con un signo positivo (+) se señala que la puntuación promedio de la entidad es superior significativamente a la media nacional. Con

un signo negativo (-) se indica que las puntuaciones de la entidad son inferiores a esta media, y las celdas vacías indican que la entidad federativa no difiere de la media nacional.

En el cuadro 8 podemos apreciar cuatro tipos de resultados: 1) entidades que presentan sólo diferencias a favor de 2003 a 2004 en una o más pruebas, 2) entidades que presentan sólo diferencias en contra, 3) entidades que no presentan diferencias y 4) entidades que presentan

CUADRO 8 • Diferencias significativas entre las medias de las entidades federativas con la media nacional, en comprensión lectora y matemáticas, de primaria y secundaria.

Entidad federativa	6° de primaria				3° de secundaria			
	Comprensión lectora		Matemáticas		Comprensión lectora		Matemáticas	
	2003	2004	2003	2004	2003	2004	2003	2004
<i>Aguascalientes</i>		+	+	+				
<i>Baja California</i>	+				+			-
<i>Baja California Sur</i>	+				+	+	+	+
<i>Campeche</i>	-	-		-			-	-
<i>Coahuila</i>	+	+		+			-	
<i>Colima</i>		+			+	+	+	+
<i>Chiapas</i>	-	-		-	-	-	-	-
<i>Chihuahua</i>	+	+						
<i>Distrito Federal</i>	+	+		+	+	+	+	+
<i>Durango</i>		-	+	-	+	+	+	+
<i>Guanajuato</i>			+					
<i>Guerrero</i>	-	-		-	-	-	-	-
<i>Hidalgo</i>								
<i>Jalisco</i>	+	+		+		+		+
<i>Estado de México</i>						-		-
<i>Michoacán</i>		-	-	-				
<i>Morelos</i>		+		+				
<i>Nayarit</i>								
<i>Nuevo León</i>	+	+						
<i>Oaxaca</i>	-	-		-	-	-		
<i>Puebla</i>	-					-	-	
<i>Querétaro</i>		+						+
<i>Quintana Roo</i>		+			+	+		
<i>San Luis Potosí</i>								
<i>Sinaloa</i>	+	+	+	+				
<i>Sonora</i>						-		-
<i>Tabasco</i>	-	-	-	-			-	-
<i>Tamaulipas</i>	+	+				+		
<i>Tlaxcala</i>		+		+				
<i>Veracruz</i>	-	-	-	-				
<i>Yucatán</i>								
<i>Zacatecas</i>	-	-			-	-		-

Fuente: INEE (2003, 2004). La calidad de la educación básica en México.

+ Desempeño medio estadísticamente significativo superior al promedio nacional (confiabilidad de 90 o más).

- Desempeño medio estadísticamente significativo inferior al promedio nacional (confiabilidad de 90 o más).

diferencias a favor en un área temática y en contra en otra. Sin embargo, para observar con mayor facilidad estas diferencias presentamos el cuadro 9 donde sólo se señalan los cambios ocurridos de 2003 a 2004, respecto a la posición de la entidad con la media poblacional. Con un signo positivo (+) se señala que hubo una mejoría en el grado y habilidad correspondiente, y con un signo negativo (-) se señala que hubo un retroceso en alguna de las pruebas. Las celdas vacías indican que no hubo cambios de un año a otro.

En el cuadro 9 queda claro que las entidades que sufrieron sólo cambios positivos fueron Aguascalientes, Coahuila, Colima, Jalisco, Morelos, Puebla, Quintana Roo, Tamaulipas y Tlaxcala. Las entidades que sufrieron cambios positivos y negativos fueron: Estado de México y Oaxaca. Las entidades que no sufrieron cambios fueron Chihuahua, Distrito Federal, Guanajuato, Hidalgo, Nuevo León, San Luis Potosí, Sinaloa, Tabasco, Veracruz y Yucatán. Finalmente, los estados que sufrieron solo cambios negativos fueron Baja California, Baja California Sur, Campeche, Chiapas, Durango, Guerrero, Michoacán, Sonora y Zacatecas.

De nuevo, nos debemos preguntar: ¿Por qué se presentan estos cambios en ausencia de alguna razón aparente? Trataremos de responder a esta pregunta en el siguiente apartado.

CONCLUSIONES

Es una práctica común que en los proyectos internacionales de evaluación educativa, como es el caso de PISA⁴ y TIMSS,⁵ los países participantes se ordenen en función de la puntuación promedio obtenida por los alumnos en cada una de las pruebas. Este sencillo procedimiento ofrece

una visión que parece clara e indiscutible sobre la calidad del sistema educativo de los países participantes. Asimismo, en los dos informes anuales que ha publicado el INEE desde su creación (2003 y 2004) se ha utilizado esta forma de presentar los resultados, comparando los distintos estratos y modalidades educativos, así como las entidades federativas del país.

Siguiendo el mismo esquema de presentación de resultados, en este trabajo nos propusimos comparar las puntuaciones de las aplicaciones de las Pruebas Nacionales de 2003 y 2004, que coordinó la DGE y el INEE, respectivamente. Sin embargo, las diferencias en los instrumentos y en los procedimientos de administración de las pruebas impusieron ciertas limitaciones a nuestro objetivo: la más importante se relaciona con la imposibilidad de comparar las puntuaciones de dos instrumentos que no son equiparables. Como consecuencia, decidimos comparar las conclusiones a las que se llega en ambos estudios.

Los resultados obtenidos muestran que en ambas aplicaciones el ordenamiento de los estratos y modalidades educativas fue el mismo en ambas habilidades. Para las escuelas de sexto grado de primaria, se observó el siguiente orden: privadas, urbanas, rurales, cursos comunitarios e indígenas. Para tercero de secundaria el ordenamiento fue el siguiente: privadas, generales, técnicas y telesecundarias. Sin embargo, este orden jerárquico no es del todo correcto si se toman en cuenta los niveles de valor estadístico (por ejemplo, entre escuelas rurales y cursos comunitarios, o entre secundarias generales y técnicas). Aún más, cuando contrastamos el valor de estas diferencias entre un año y otro encontramos discrepancias que muestran la inestabilidad de los resultados.

CUADRO 9 • Cambios significativos de 2003 a 2004 en las entidades federativas respecto a la media nacional.				
Entidad federativa	6° de primaria		3° de secundaria	
	Comprensión lectora	Matemáticas	Comprensión lectora	Matemáticas
Aguascalientes	+			
Baja California	-		-	-
Baja California Sur	-			
Campeche		-		
Coahuila		+		+
Colima	+			
Chiapas		-		
Chihuahua				
Distrito Federal				
Durango	-	-		
Guanajuato				
Guerrero		-		
Hidalgo				
Jalisco		+	+	+
Estado de México		+	-	-
Michoacán	-	-		
Morelos	+	+		
Nayarit				
Nuevo León				
Oaxaca		-		+
Puebla	+		-	
Querétaro	+			+
Quintana Roo	+			
San Luis Potosí				
Sinaloa				
Sonora			-	-
Tabasco				
Tamaulipas			+	
Tlaxcala	+	+		
Veracruz				
Yucatán				
Zacatecas				-

Por otro lado, cuando analizamos los resultados de las entidades federativas de un año a otro, encontramos que algunas de ellas permanecen sin cambio, otras mejoran o empeoran en algunas habilidades y niveles educativos, mientras que otras mejoran en algunos aspectos, a la vez que empeoran en otros. Pero, ¿será posible que en el lapso de un año pueda cambiar el aprovechamiento de los estudiantes, en ausencia de reformas educativas de envergadura? ¿No ocurrirá que los cambios observados sean un artificio de los instrumentos y procedimientos utilizados? Analicemos esta hipótesis.

En primer lugar, debemos decir que las pruebas de aprendizaje evalúan, en palabras de James Popham (1990), tres elementos de los estudiantes que se entremezclan de manera indivisible: 1) sus capacidades intelectuales, 2) lo que aprenden en el hogar y en su contexto social y 3) lo que aprenden en la escuela. Así, los resultados de aprendizaje son bastante congruentes con el sentido común: las escuelas cuyos alumnos tengan las mejores condiciones personales, sociales y escolares tendrán los mejores resultados. De aquí que no sea una sorpresa que las escuelas privadas obtengan las puntuaciones más altas y que las escuelas indígenas y las telesecundarias obtengan las calificaciones más bajas.

El mismo razonamiento se aplica a los resultados de las entidades federativas. Es lógico que el Distrito Federal, Jalisco, Colima y Aguascalientes estén siempre por encima de la media nacional, y que los estados de Chiapas, Oaxaca, Guerrero y Veracruz se encuentren por debajo. Sólo la cantidad de escuelas rurales, indígenas y telesecundarias de estos estados explican en gran parte sus posiciones en el cuadro de los *rankings* estatales. Tanto las evaluaciones nacionales como las internaciona-

les son congruentes con esta hipótesis; lo que no impide que haya sus notables excepciones.

Sin embargo, los argumentos anteriores no explican por qué las entidades federativas cambian de un año a otro. La hipótesis más plausible se encuentra en la diferencia de los instrumentos evaluativos, así como en posibles fallas de los procedimientos utilizados. Al inicio del trabajo explicamos la no equiparabilidad de las dos pruebas, razón que explica en mucho la diferencia de los resultados obtenidos.

Una razón que ayuda a explicar estas incongruencias radica en la forma en cómo se seleccionó a los estudiantes en ambos estudios. Como se mencionó con anterioridad, las evaluaciones nacionales, desde 1998, partieron de un mismo marco muestral y las muestras poblacionales se diseñaron con una buena metodología. Sin embargo, en vez de utilizar distintas escuelas cada año, se utilizaron las mismas escuelas en todos los casos; condición que pudo contaminar el proceso de aplicación de las pruebas y, con ello, alterar los resultados obtenidos. A esta condición podemos agregar otra: que el INEE, por primera ocasión, coordinó la aplicación nacional de 2004, lo cual nos hace suponer que la actitud de las escuelas hacia la evaluación pudo haber cambiado.

En síntesis, los resultados muestran que las evaluaciones nacionales realizadas en 2003 y 2004 son sensibles y relativamente consistentes para discriminar grandes poblaciones, como son los estratos y modalidades educativos. Sin embargo, son menos sensibles y más inconsistentes cuando las poblaciones se reducen, como es el caso de las entidades federativas. Por último, es de esperarse que fueran poco

sensibles y muy inconsistentes para discriminar poblaciones más pequeñas, como los estratos y las modalidades educativos de cada entidad federativa; ni qué decir de las escuelas en lo individual. De haber *rankings* en este nivel de desagregación, no nos extrañaría que éstos cambiaran constantemente de un año a otro, sin una razón aparente.

Finalmente, una pregunta obligada en un estudio comparativo como éste es si el sistema educativo nacional ha avanzado o retrocedido en el tiempo. Desafortunadamente, analizando los resultados de ambas evaluaciones, no podemos concluir una cosa u otra pues, además de que los instrumentos no son equiparables, la implementación del muestreo no fue la más afortunada; razón por la cual los resultados muestran inconsistencias que no se pueden atribuir a cambios reales del sistema educativo nacional.

Por fortuna, el INEE ha iniciado diversas acciones orientadas a dilucidar las tendencias educativas del país. Por un lado, acaba de publicar un reporte de los resultados mexicanos en PISA (Vidal y Díaz, 2004), comparando los estudios de 2000 y 2003. Por otro lado, está preparando en este año una réplica de las Pruebas de Estándares Nacionales utilizadas en 2000, con el fin de poder determinar el avance del sistema educativo nacional en el último lustro.⁶ Finalmente, está desarrollando una nueva generación de Exámenes de la Calidad y el Logro Educativos (EXCALE) que permitan conocer lo que los estudiantes aprenden a lo largo del tiem-

po. Esperemos que con estas acciones evaluativas tengamos una mejor comprensión del sistema educativo nacional en el tiempo.

REFERENCIAS

- ARCE-Ferrer, A. y E. Backhoff (2002), "Efecto del método de igualación en la equivalencia de los exámenes nacionales", en *Revista Mexicana de Psicología*, vol. 19, núm. 2, pp. 101-114.
- BACKHOFF, E. (2004), "La comparación entre entidades: alcances y limitaciones de los *rankings*", en *Memorias de las Jornadas de Evaluación Educativa*, México.
- BACKHOFF, E. y F. Martínez-Rizo (2004), "Resultados de las Pruebas de Estándares Nacionales 2003: elementos para la comparación entre entidades", en *Memorias del VI Foro de Evaluación Educativa*, Aguascalientes.
- INSTITUTO Nacional para la Evaluación de la Educación (INEE) (2004), *La calidad de la educación básica en México: resultados de evaluación educativa 2004*, México, INEE.
- (2003), *La calidad de la educación básica en México: primer informe anual 2003*, México, INEE.
- MARTÍNEZ-RIZO, F. (2004), "Compatibilidad de los resultados de las evaluaciones", en *Memorias de las Jornadas de Evaluación Educativa*.
- PODER EJECUTIVO (2002), Decreto de creación del Instituto Nacional para la Evaluación de la Educación, México, *Diario Oficial*, 8 de agosto.
- VIDAL y M. A. Díaz (2004), "Resultados de las Pruebas PISA 2002 y 2003 en México, Habilidades para la vida en estudiantes de 15 años", México, INEE.

NOTAS

1. Reactivos comunes que comparten dos pruebas cuyos resultados se desean igualar.
2. Proceso estadístico para hacer equivalentes las puntuaciones de dos pruebas (Arce-Ferrer y Backhoff, 2002).
3. Medida que indica cuán consistentes son las puntuaciones de una prueba.
4. Por sus siglas en inglés: Programa Internacional para la Evaluación de los Estudiantes.
5. Por sus siglas en inglés: Tercer Estudio Internacional de Matemáticas y Ciencias Naturales.
6. Con las limitaciones propias de los instrumentos y procedimientos utilizados.