

Informe Técnico de la Evaluación NAEP de Matemáticas en Puerto Rico

ENFOQUE EN ASUNTOS ESTADÍSTICOS

Evaluación Nacional del Progreso Educativo

Índice

Diciembre de 2007

- 1 Resumen Ejecutivo
- 2 Administraciones de 2003 y 2005 en Puerto Rico
- 10 El Desajuste de Preguntas
- 14 Informe de los Resultados de Puerto Rico en la Escala de NAEP
- 22 Integración de Puerto Rico a NAEP: Próximos Pasos
- 24 Muestreo e Implementación
- 26 Referencias
- 27 Apéndice

¿Qué es la Libreta de Calificaciones de la Nación™?

La Libreta de Calificaciones de la Nación informa al público sobre el logro académico de los estudiantes de escuelas elementales y secundarias en los Estados Unidos y sus jurisdicciones, incluyendo a Puerto Rico. Las libretas de calificaciones comunican los hallazgos de la Evaluación Nacional del Progreso Educativo (NAEP, por sus siglas en inglés), una medida continua y nacionalmente representativa de los logros en varias materias a lo largo del tiempo. La Libreta de Calificaciones de la Nación compara el rendimiento entre estados, distritos urbanos, escuelas públicas y privadas y grupos demográficos de estudiantes.

Por más de tres décadas, las evaluaciones NAEP se han llevado a cabo periódicamente en lectura, matemáticas, ciencias, escritura, historia, geografía y otras materias. Al proporcionar información objetiva sobre el rendimiento estudiantil a nivel nacional, estatal y local, NAEP es una parte esencial de la evaluación nacional de la condición y el progreso de la educación.

Sólo se recopila información relacionada con el logro académico y las variables pertinentes. Se protege la privacidad de los estudiantes individuales y no se revelan las identidades de las escuelas que participan en la evaluación.

NAEP es un proyecto del Centro Nacional para Estadísticas de la Educación (NCES, por sus siglas en inglés) que se lleva a cabo por mandato del Congreso, dentro del Instituto de Ciencias de la Educación del Departamento de Educación de los Estados Unidos. El Comisionado de Estadísticas de Educación es responsable de llevar a cabo el proyecto NAEP. La Junta Regidora de la Evaluación Nacional (National Assessment Governing Board) supervisa y establece la política para NAEP.

Resumen Ejecutivo

En 2003 la evaluación NAEP de matemáticas se administró en español, a modo de prueba, a estudiantes de escuelas públicas de cuarto y octavo grado en Puerto Rico. A partir de análisis preliminares de los datos de 2003, se realizaron cambios a los procedimientos de administración y traducción para la administración de NAEP de 2005 en Puerto Rico. Este informe describe el contenido y la administración de la evaluación NAEP de matemáticas en Puerto Rico a modo de prueba en 2003 y 2005, los problemas con el desajuste de preguntas en los datos de 2003, los resultados de un análisis especial de validez y los planes para integrar a Puerto Rico a la muestra nacional en futuras administraciones.

La evaluación NAEP de matemáticas a modo de prueba en Puerto Rico fue administrada en español. Análisis preliminares de los datos de 2003 para Puerto Rico suscitaron inquietudes de que las preguntas no estaban funcionando como en otras jurisdicciones. En comparación con otras jurisdicciones, en Puerto Rico hubo una mayor cantidad de datos faltantes, menos respuestas correctas de lo esperado para cada área de contenido y una mayor falta de concordancia entre el rendimiento esperado y el rendimiento real de los estudiantes en las preguntas (desajuste de preguntas).

Para mejorar la calidad de los datos, se realizaron modificaciones a los procedimientos de traducción y administración para la evaluación de 2005 en Puerto Rico. Estos cambios incluyeron la revisión del manual del administrador, el otorgamiento de 10 minutos adicionales para cada una de las dos secciones cronometradas de la evaluación y la ampliación de los procedimientos de traducción. El análisis de los datos de 2005 mostró una menor cantidad de respuestas faltantes y un mayor porcentaje de respuestas correctas en comparación con los de 2003. A pesar de estas mejoras en la calidad de los datos, hubo inquietud sobre la validez de informar los resultados de Puerto Rico en la escala de NAEP.

Para afrontar esta inquietud, se realizó un estudio de validez con los resultados de 2003 y 2005 de Puerto Rico. El análisis involucró eliminar preguntas que mostraron desajuste en la muestra de Puerto Rico,

dando lugar así a una escala restringida. Luego, el rendimiento en el conjunto reducido de preguntas (escala restringida) se comparó con el rendimiento correspondiente en el conjunto total de preguntas (escala completa) para Estados Unidos y para Puerto Rico. Las dos escalas coincidieron dentro de tres décimas de un punto de escala para Estados Unidos, pero hubo una diferencia de 2 a 3 puntos entre las escalas para Puerto Rico. Sin embargo, esta diferencia no cambia el *ranking* de Puerto Rico con respecto a otras jurisdicciones participantes. Estos hallazgos indican que los resultados para Puerto Rico se pueden informar en la escala de NAEP de 0 a 500.

A partir de 2003, se requiere que todas las jurisdicciones que reciben fondos federales de Título I, incluyendo a Puerto Rico, participen en NAEP en cuarto y octavo grado en años alternos. Las administraciones de 2003 y 2005 en Puerto Rico se consideraron a modo de prueba, y los resultados no se informaron con los de otras jurisdicciones. En futuras administraciones de NAEP, la intención es incluir a Puerto Rico dentro de la muestra de Estados Unidos. Para la administración de 2007, NCES ha incrementado la participación de educadores de Puerto Rico en el proceso de desarrollo y revisión de la traducción de la evaluación NAEP de matemáticas. Se han implementado acciones para encaminar a Puerto Rico hacia una total integración al muestreo, la recopilación de datos y el informe para la administración de 2009.

Respecto a este informe

Este informe pertenece a una serie de tres informes sobre la administración y los resultados de las evaluaciones NAEP de matemáticas a modo de prueba de 2003 y 2005 en Puerto Rico, disponibles en http://nationsreportcard.gov/puertorico_2005/. El primer informe, *Matemáticas 2003 y 2005, Rendimiento en Puerto Rico: Aspectos Sobresalientes*, presenta resultados para Puerto Rico y para Estados Unidos en términos de puntuaciones en escala y niveles de logro de NAEP. El segundo informe, *Matemáticas 2005, Rendimiento en Puerto Rico: Enfoque en las Áreas de Contenido*, proporciona resultados por área de contenido e incluye un análisis sobre el rendimiento de los estudiantes en ejemplos de preguntas. Éste, el tercer informe, se centra en consideraciones técnicas de las evaluaciones a modo de prueba y en los planes para incluir a Puerto Rico como parte de la muestra de Estados Unidos en futuras administraciones.



Capítulo 1

Administraciones de 2003 y 2005 en Puerto Rico

La evaluación NAEP de matemáticas se tradujo al español para permitir que Puerto Rico participase, a modo de prueba, en la administración de 2003. En 2005 la evaluación NAEP de matemáticas se administró nuevamente a estudiantes de cuarto y octavo grado de escuelas públicas en Puerto Rico. Un objetivo fundamental de las administraciones a modo de prueba era informar sobre los resultados de Puerto Rico en la escala de Estados Unidos. Este capítulo describe las administraciones de NAEP de matemáticas de 2003 y 2005 en Puerto Rico, las inquietudes sobre la calidad de los datos que surgieron en 2003 y los cambios realizados en 2005 para afrontar dichas inquietudes.

Evaluación NAEP de Matemáticas de 2003 en Puerto Rico

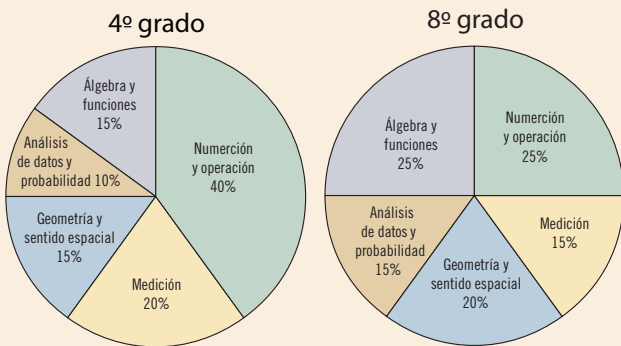
El Título I de la Ley de Educación Elemental y Secundaria de 1965, según enmendado, exige que todas las jurisdicciones que reciben fondos federales de Título I participen, en cuarto y octavo grado, en NAEP de lectura y matemáticas en años alternos a partir de 2003. El Departamento de Educación de los Estados Unidos decidió que Puerto Rico no debería participar en NAEP de lectura, ya que esa evaluación mide la capacidad de los estudiantes para leer en inglés, y el español es el idioma de instrucción en Puerto Rico.

La evaluación NAEP de matemáticas de 2003 se tradujo al español para permitir que Puerto Rico participase a modo de prueba.

El objetivo principal de la administración a modo de prueba de 2003 en Puerto Rico fue administrar la evaluación NAEP de matemáticas utilizando los mismos procedimientos utilizados en otras jurisdicciones, ya que en el futuro los resultados de la evaluación en Puerto Rico se han de comparar a lo largo del tiempo y con los de otras jurisdicciones. La administración de NAEP en Puerto Rico fue consistente con la de otras jurisdicciones en lo que respecta al contenido, tipos de preguntas, muestreo, administración y procedimientos de calificación.

Contenido. El contenido de la evaluación de matemáticas se basa en un marco teórico que describe en detalle cómo NAEP debe evaluar las matemáticas. El marco teórico de matemáticas de NAEP especifica el contenido a evaluar en cada grado y el porcentaje de preguntas a evaluar en cada una de cinco áreas de contenido. En 2003 las cinco áreas de contenido fueron (1) numeración y operación; (2) medición; (3) geometría y sentido espacial; (4) análisis de datos y probabilidad; y (5) álgebra y funciones. La distribución porcentual de las preguntas para cuarto y octavo grado se muestra en la Figura 1.1.

Figura 1.1
Distribución porcentual de las preguntas NAEP de matemáticas, por grado y área de contenido, en 2003



FUENTE: Marco teórico de matemáticas para la Evaluación Nacional del Progreso Educativo de 2003, Junta Regidora de la Evaluación Nacional, Departamento de Educación de los EE. UU., septiembre de 2002.

Además del contenido, el marco teórico especificaba que cada pregunta debía medir una de tres habilidades matemáticas: comprensión conceptual, conocimiento de procedimientos y resolución de problemas. Los marcos teóricos están disponibles (en inglés) en <http://www.nagb.org/pubs/pubs.html>.

Tipos de preguntas. La evaluación NAEP de matemáticas incluye una combinación de preguntas de selección múltiple y de respuesta construida. Las preguntas de selección múltiple requieren que los estudiantes seleccionen la respuesta correcta de entre

cuatro o cinco posibles opciones. Las respuestas a estas preguntas se califican como correctas o incorrectas. Las preguntas de respuesta construida corta requieren que los estudiantes proporcionen respuestas a problemas de cálculo o que describan soluciones en una o dos oraciones. Las preguntas de respuesta construida extendida requieren que los estudiantes proporcionen respuestas escritas de más de una o dos oraciones. Estas preguntas están diseñadas para medir la habilidad de los estudiantes para razonar, comunicar y establecer conexiones entre conceptos y destrezas, ya sea entre las cinco áreas de contenido de matemáticas o de las matemáticas a otras áreas curriculares. Las respuestas a las preguntas de respuesta construida se califican como correctas o incorrectas, o se califican con uno o más niveles de crédito parcial.

Traducción. Primeramente, para el acomodo bilingüe inglés / español de 1996, un subconjunto de materiales de matemáticas de NAEP fue traducido por un equipo de habla hispana de expertos en evaluación. Se consultaron libros de texto de matemáticas utilizados en educación bilingüe para garantizar que los contextos fuesen plausibles y que la terminología matemática fuese correcta. Un panel de educadores de habla hispana que representaba múltiples variantes lingüísticas del idioma, incluyendo español de Puerto Rico, evaluó la versión en español de las preguntas.

Previo a 2004, el enfoque del proceso de traducción/adaptación se centró en el desarrollo de una única versión en español que pudiera administrarse universalmente a estudiantes de origen puertorriqueño, cubano, mexicano, centroamericano u otro origen de habla hispana. Más que una traducción literal del inglés al español¹, el objetivo fue producir una evaluación psicométricamente equivalente. Tras las revisiones del panel, especialistas bilingües en lengua realizaron revisiones editoriales y de equidad/imparcialidad de todas las preguntas.

¹ Ver consideraciones acerca de la realización de evaluaciones en múltiples idiomas en Hambleton, RK., Merenda, P. F., y Spielberger, C.D. (2005).



Muestreo y administración. En Puerto Rico, para la administración a modo de prueba de NAEP de 2003, se muestrearon aproximadamente 100 escuelas públicas y 3,000 estudiantes de cada uno de los dos grados, cuarto y octavo. Los detalles de los procedimientos de muestreo y las tasas de participación pueden encontrarse en la página 24–25 de este informe.

De conformidad con los procedimientos utilizados en otras jurisdicciones, el periodo de administración comenzó la última semana de enero y continuó hasta la primera semana de marzo de 2003. Cada estudiante recibió un folleto que contenía dos bloques de preguntas a completarse en 25 minutos, para un tiempo de evaluación total de 50 minutos. Además de las preguntas de matemáticas, se les pidió a los estudiantes que respondieran a preguntas sobre su trasfondo académico, incluyendo preguntas específicas a su instrucción y experiencias en matemáticas.

Las sesiones de evaluación fueron llevadas a cabo por administradores que no eran miembros del personal de las escuelas participantes. Estos administradores fueron contratados y capacitados en Puerto Rico para minimizar el impacto de diferencias lingüísticas y culturales en la administración. Los materiales de los administradores, incluyendo manuales e instrucciones, fueron traducidos al español.

Calificación. Los resultados de NAEP de matemáticas se informan de dos maneras: como puntuaciones promedio en la escala de NAEP de matemáticas y como porcentajes de estudiantes que alcanzan distintos

niveles de logro de NAEP en matemáticas. Los resultados del rendimiento general y del rendimiento en cada una de las cinco áreas de contenido se informan en una escala de 0 a 500. Las puntuaciones en escala se calculan para grupos de estudiantes, no para estudiantes individuales.

Además de las puntuaciones en escala, los resultados se presentan en términos de los niveles de logro en matemáticas adoptados por la Junta Regidora. Los niveles de logro están destinados a medir cuán bien los logros reales de los estudiantes corresponden con los logros que se esperan de ellos. Para cada grado evaluado, la Junta Regidora ha adoptado tres niveles de logro: Básico, Competente y Avanzado. La información sobre cómo la Junta Regidora establece los niveles de logro puede encontrarse (en inglés) en <http://nces.ed.gov/nationsreport-card//pubs/main1996/97951.asp>.

Niveles de Logro de NAEP

Los tres niveles de logro de NAEP, del más bajo al más alto, son *Básico*, *Competente* y *Avanzado*.

Básico: Este nivel denota dominio parcial de conocimientos y destrezas, que son requisitos previos, fundamentales para el rendimiento competente en cada grado.

Competente: Este nivel representa un rendimiento académico sólido para cada grado evaluado. Los estudiantes que alcanzan este nivel han demostrado competencia en contenido exigente de la materia, incluyendo conocimiento de la materia, aplicación de dicho conocimiento a situaciones de la vida real y las destrezas analíticas adecuadas para la materia.

Avanzado: Este nivel significa rendimiento superior.

Puntuaciones de corte

Las puntuaciones de corte representan la puntuación mínima requerida para el rendimiento en cada nivel de logro en la evaluación NAEP. Las puntuaciones de corte de matemáticas en la escala de NAEP de 0 - 500 que definen los límites inferiores de cada uno de los niveles de logro son:

	4º Grado	8º Grado
<i>Básico</i>	214	262
<i>Competente</i>	249	299
<i>Avanzado</i>	282	333

Análisis Preliminar de la Administración de 2003

La administración NAEP de 2003 en Puerto Rico fue el primer intento de llevar a cabo una administración completa en un idioma distinto al inglés. El análisis preliminar de los datos para la evaluación de matemáticas de 2003 indicó tres diferencias importantes entre Puerto Rico y otras jurisdicciones. En comparación con otras jurisdicciones, en Puerto Rico hubo un porcentaje más elevado de respuestas faltantes, un porcentaje más elevado de respuestas incorrectas y niveles más elevados de desajuste de preguntas.

Porcentaje elevado de datos faltantes. La Tabla 1.1 muestra el porcentaje de respuestas faltantes para Puerto Rico y Estados Unidos en 2003. Las respuestas faltantes incluyen las preguntas omitidas y las preguntas no alcanzadas. Una pregunta fue considerada omitida si un estudiante la saltó pero respondió una o más de las preguntas siguientes. Una pregunta fue considerada no alcanzada cuando ni esa pregunta ni

ninguna de las preguntas siguientes en esa sección fue contestada. Para la evaluación de 2003 en su totalidad, en promedio, un 25 por ciento de los estudiantes de cuarto grado en Puerto Rico dejó cada pregunta sin contestar, comparado con un 7 por ciento en Estados Unidos. En 2003, para los estudiantes de octavo grado, los porcentajes respectivos de respuestas faltantes fueron 19 por ciento en Puerto Rico y 5 por ciento en Estados Unidos. En todas las áreas de contenido en ambos grados hubo más respuestas faltantes en Puerto Rico que en Estados Unidos.

El porcentaje de respuestas faltantes en Puerto Rico fue más elevado que el de Estados Unidos para todos los tipos de pregunta en ambos grados. En promedio, para las preguntas de respuesta construida corta, el porcentaje de respuestas faltantes fue de 39 por ciento en cuarto grado y de 29 por ciento en octavo grado. Para Estados Unidos, los porcentajes correspondientes de respuestas faltantes a preguntas de respuesta construida corta fueron de 9 por ciento y 6 por ciento en cuarto y octavo grado, respectivamente.

Tabla 1.1

Porcentaje promedio de respuestas faltantes en la evaluación NAEP de matemáticas para estudiantes de escuelas públicas en Puerto Rico y Estados Unidos de 4º y 8º grado, por área de contenido y tipo de pregunta, en 2003

Característica de la pregunta	4º grado			8º grado		
	Número de preguntas	Porcentaje de respuestas faltantes		Número de preguntas	Porcentaje de respuestas faltantes	
		Puerto Rico	EE. UU.		Puerto Rico	EE. UU.
Total	179	25*	7	195	19*	5
Área de contenido						
Numeración y operación	75	23*	6	51	14*	4
Medición	32	21*	6	30	16*	4
Geometría	27	25*	6	36	22*	5
Análisis de datos y probabilidad	19	28*	5	29	24*	7
Álgebra y funciones	26	32*	10	49	20*	6
Tipo de pregunta						
Selección múltiple	114	16*	4	126	11*	4
Respuesta construida corta	57	39*	9	60	29*	6
Respuesta construida extendida	8	59*	22	9	62*	25

* Puerto Rico significativamente diferente de Estados Unidos ($p < .05$).

NOTA: Las respuestas faltantes incluyen las preguntas omitidas y no alcanzadas.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003.

Porcentaje más elevado de respuestas incorrectas.

Además del alto porcentaje de respuestas faltantes, también hubo un porcentaje más alto de respuestas incorrectas en Puerto Rico comparado con Estados Unidos. Este hallazgo podría indicar un problema con la forma en que funcionaron las preguntas en Puerto Rico, o podría reflejar diferencias entre el conocimiento y las habilidades de los estudiantes de Puerto Rico y los de los estudiantes en Estados Unidos. La Tabla 1.2 presenta el promedio de la media de las puntuaciones de las preguntas en general, por área de contenido de matemáticas y por tipo de pregunta. Para preguntas de selección múltiple y de respuesta construida corta que se califican dicotómicamente (correctas o incorrectas), la puntuación promedio de las preguntas refleja la proporción de respuestas correctas para una pregunta en particular.

Todas las preguntas de respuesta construida extendida y algunas de las preguntas de respuesta construida corta se califican utilizando guías de evaluación de múltiples niveles. Los estudiantes pueden obtener crédito con una respuesta parcialmente correcta.

Para preguntas como éstas, la puntuación promedio de las preguntas se define como la proporción promedio

del máximo de la puntuación obtenida. Los datos de la Tabla 1.2 presentan la media para todas las preguntas.

En general, la media de la puntuación promedio de las preguntas para Puerto Rico fue de 0.27 en cuarto y octavo grado, comparada con 0.54 para Estados Unidos. Para cada una de las cinco áreas de contenido de la evaluación NAEP de matemáticas de cuarto y octavo grado, la media de las puntuaciones promedio de las preguntas fue inferior en Puerto Rico que en Estados Unidos. Se observó el mismo patrón de resultados para cada tipo de pregunta.

Niveles elevados de desajuste de preguntas. El resultado más problemático de la evaluación de matemáticas de Puerto Rico en 2003 fue la incidencia relativamente alta de desajuste de preguntas. El desajuste de pregunta se define como la falta de concordancia entre el rendimiento esperado y el rendimiento real de los estudiantes en una pregunta. Aunque las preguntas funcionaron como se esperaba en Estados Unidos, las preguntas traducidas no funcionaron como se esperaba en Puerto Rico. El capítulo 2 presenta una explicación del desajuste de preguntas.

Tabla 1.2

Diferencia entre la media de las puntuaciones promedio en la evaluación NAEP de matemáticas para estudiantes de escuelas públicas de 4º y 8º grado en Puerto Rico y Estados Unidos, por área de contenido y tipo de pregunta, en 2003

Característica de la pregunta	4º grado			8º grado		
	Puerto Rico	EE. UU.	Diferencia	Puerto Rico	EE. UU.	Diferencia
Total	.27*	.54	.27	.27*	.54	.27
Área de contenido						
Numeración y operación	.26*	.55	.29	.31*	.59	.28
Medición	.30*	.52	.22	.25*	.50	.25
Geometría	.32*	.53	.21	.29*	.55	.26
Análisis de datos y probabilidad	.24*	.59	.35	.22*	.50	.28
Álgebra y funciones	.20*	.50	.31	.26*	.55	.29
Tipo de pregunta						
Selección múltiple	.34*	.58	.24	.30*	.57	.27
Respuesta construida corta	.15*	.50	.34	.24*	.54	.31
Respuesta construida extendida	.05*	.29	.24	.06*	.25	.19

* Puerto Rico significativamente diferente de Estados Unidos ($p < .05$).

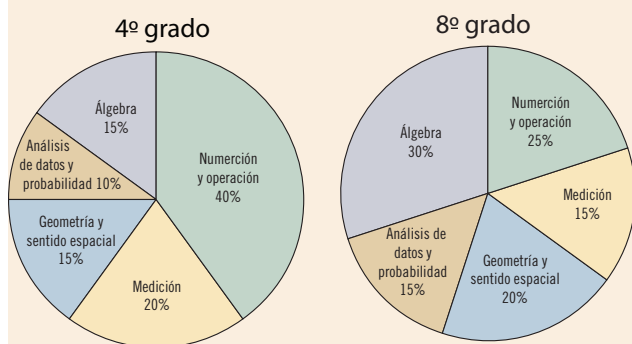
NOTA: La suma de las cifras puede no corresponder al total debido al redondeo. Las diferencias se basan en estimados no redondeados.

FUENTE: Departamento de Educación de los EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003.

Evaluación NAEP de Matemáticas de 2005 en Puerto Rico

La evaluación NAEP de matemáticas se administró a modo de prueba por segunda vez en Puerto Rico en 2005. Los detalles de los procedimientos de muestreo y las tasas de participación pueden encontrarse en la página 24–25 de este informe. Al igual que en 2003, se muestrearon aproximadamente 100 escuelas públicas y 3,000 estudiantes en cada uno de los dos grados, cuarto y octavo. Tanto en Estados Unidos como en Puerto Rico, hubo un cambio en el contenido de la evaluación para reflejar el cambiante énfasis curricular y los objetivos especificados en el marco teórico de NAEP de matemáticas. En 2005, en inglés, se modificaron los nombres de algunas áreas de contenido, pero el porcentaje de preguntas a evaluar en cada una de las cinco áreas de contenido siguió siendo el mismo para cuarto grado (Figura 1.2). Para octavo grado hubo una reducción del 5 por ciento en el énfasis en numeración y operación y un correspondiente aumento del 5 por ciento en el énfasis en álgebra comparado con el 2003.

Figura 1.2
Distribución de las preguntas NAEP de matemáticas, por grado y área de contenido, en 2005



FUENTE: Marco teórico de matemáticas para la Evaluación Nacional del Progreso Educativo de 2005, Junta Regidora de la Evaluación Nacional, Departamento de Educación de EE. UU., septiembre de 2004.

Además de especificar el contenido, el marco teórico para 2005 requiere una evaluación que mida diferentes niveles de complejidad matemática. Cada nivel de complejidad matemática (elevado, moderado y bajo) incluye aspectos de saber y hacer matemáticas, tales como razonar, realizar procedimientos, entender conceptos y resolver problemas. El nivel de complejidad de

una pregunta se determina por las exigencias cognitivas que la misma plantea a los estudiantes. Por ejemplo, una pregunta con un nivel de complejidad elevado para cuarto grado puede pedir a los estudiantes que expliquen y justifiquen sus soluciones a un problema.

Cambios a la Administración de 2005 en Puerto Rico

Las inquietudes en cuanto a la calidad de los datos de Puerto Rico en 2003 condujeron a tres cambios para NAEP de 2005 en Puerto Rico a modo de prueba. Estos cambios incluyeron el perfeccionamiento de los procedimientos de traducción, la revisión de los procedimientos de administración y la adición de talleres para desarrollar comprensión de NAEP y para fomentar la participación en NAEP.

Procedimientos perfeccionados de traducción. Una compañía de traducción contratada por el Centro Nacional para Estadísticas de la Educación (NCES) realizó una revisión independiente de verificación de la traducción. Estos procedimientos de revisión de verificación de la traducción se llevaron a cabo después de que los folletos de evaluación de 2003 estuvieron impresos y por lo tanto los cambios generados por este proceso se aplicaron por primera vez para la evaluación de 2005. Los procedimientos de traducción se perfeccionaron de otras tres maneras. Primero, dos maestros de matemáticas de Puerto Rico se incorporaron al panel de revisión de expertos. Estos dos maestros de cuarto y octavo grado de Puerto Rico brindaron una valiosa perspectiva sobre cómo los estudiantes en Puerto Rico podrían interpretar palabras específicas y responder a ciertos contextos. Segundo, algunos contextos y léxico utilizados en las preguntas de matemáticas fueron adaptados a las características lingüísticas y culturales particulares de Puerto Rico. Tercero, todo el conjunto de cuestionarios de trasfondo se tradujo y se adaptó para usarse en Puerto Rico utilizando los mismos parámetros que guiaron la traducción de las preguntas de la evaluación.

Cambios en procedimientos de administración. Se implementaron dos cambios en el proceso de administración. En primer lugar, se revisó el guión de la administración en Puerto Rico para dar a los estudiantes instrucciones explícitas de pasar a la pregunta

siguiente en vez de continuar con una pregunta para la cual no supieran la respuesta. El guión revisado también incluyó amplias explicaciones sobre los diferentes tipos de preguntas en la evaluación. Estos cambios se hicieron porque era posible que los estudiantes en Puerto Rico no estuvieran familiarizados con el formato de evaluación de NAEP.

En segundo lugar, en un intento por reducir la prevalencia de respuestas faltantes, se les brindaron a los estudiantes 10 minutos más para completar cada una de las dos secciones cronometradas de matemáticas. En otras jurisdicciones participantes, a los estudiantes se les brindan 50 minutos para completar la evaluación y a los estudiantes que reciben acomodos se les brindan 10 minutos más por sección para un total de 70 minutos. Así, en 2005 a los estudiantes en Puerto Rico se les brindó la misma cantidad de tiempo de evaluación que a los estudiantes que reciben acomodos en otras jurisdicciones.

Talleres de motivación. Unas semanas antes de la administración de NAEP de 2005 se realizaron dos talleres de motivación en San Juan y en Ponce. El objetivo de los talleres fue incrementar el interés en NAEP y ayudar a los funcionarios a aprender a motivar a maestros y estudiantes a participar en la administración de NAEP de 2005. En los talleres participaron cerca de 150 directores y otros funcionarios de escuelas seleccionadas para la evaluación de 2005.

Los talleres presentaron el contenido, propósito y finalidad de los informes de NAEP. Los participantes recibieron ejemplos de preguntas y manipulativos que se utilizaron en evaluaciones previas, y juntos exploraron estrategias para ayudar a los estudiantes a tener un buen desempeño en evaluaciones tales como NAEP. Los talleres fueron llevados a cabo por un equipo de habla hispana de expertos en evaluación de NAEP y contaron con la participación de representantes del Departamento de Educación de Puerto Rico.

Análisis Preliminar de la Administración de 2005

El análisis de los datos de 2005 de Puerto Rico indicó una disminución en el porcentaje de respuestas fal-

tantes y un incremento en el porcentaje de respuestas correctas comparado con 2003. Los cambios podrían deberse a las modificaciones en los procedimientos de traducción y administración descritos arriba, a un incremento en el rendimiento de los estudiantes en cuarto y octavo grado o a ambos factores.

La Tabla 1.3 muestra la reducción en el porcentaje de respuestas faltantes entre 2003 y 2005 en cuarto y octavo grado. En este análisis se utilizaron preguntas comunes a los dos años de evaluación (2003 y 2005). Para cada área de contenido y para cada tipo de pregunta y grado, el porcentaje promedio de respuestas faltantes en 2005 fue inferior al de 2003 (Tabla 1.3). Sin embargo, la cantidad de respuestas faltantes sigue siendo superior en Puerto Rico que en Estados Unidos.

La Tabla 1.4 muestra la media de puntuaciones promedio de las preguntas en 2003 y 2005 por grado y tipo de pregunta. En general y en algunas áreas de contenido hubo mejoras significativas en cuarto y octavo grado. A excepción de las preguntas de selección múltiple de cuarto grado, la media de puntuaciones promedio se incrementó de 2003 a 2005 para todos los tipos de preguntas en cada nivel de grado. A pesar de estas mejoras, siguen habiendo grandes diferencias promedio en las preguntas entre Puerto Rico y Estados Unidos.

Los cambios realizados entre las administraciones de 2003 y 2005 iban dirigidos a abordar las inquietudes sobre la calidad de los datos de 2003. Los resultados de la administración de 2005 mostraron puntuaciones promedio más elevadas y porcentajes más bajos de datos faltantes comparado con 2003. Sin embargo, las puntuaciones de la evaluación en Puerto Rico continuaron siendo significativamente más bajas que las del resto de Estados Unidos. Además, en Puerto Rico hubo más respuestas faltantes y puntuaciones promedio más bajas que en Estados Unidos. Pueden encontrarse detalles del rendimiento de los estudiantes en Puerto Rico en *Matemáticas 2003 y 2005, Rendimiento en Puerto Rico: Aspectos Sobresalientes*, disponible en http://nationsreportcard.gov/puertorico_2005/. Como se describe en el capítulo 3 de este informe, se realizaron análisis adicionales para abordar las inquietudes sobre el grado en que las preguntas traducidas utilizadas en Puerto Rico no funcionaron como se esperaba.

Tabla 1.3

Diferencia entre el promedio del porcentaje de respuestas faltantes en evaluaciones NAEP de matemáticas de 2003 y 2005 para estudiantes de 4^o y 8^o grado de escuelas públicas en Puerto Rico, por área de contenido y tipo de pregunta

Característica de la pregunta	4 ^o grado			8 ^o grado		
	Porcentaje de respuestas faltantes			Porcentaje de respuestas faltantes		
	2003	2005	Diferencia	2003	2005	Diferencia
Total¹	25*	11	14	18*	9	9
Área de contenido						
Numeración y operación	24*	10	13	15*	8	7
Medición	22*	9	13	17*	8	8
Geometría	25*	13	12	25*	13	11
Análisis de datos y probabilidad	30*	14	16	20*	11	9
Álgebra y funciones	26*	12	14	16*	6	9
Tipo de pregunta						
Selección múltiple	16*	6	10	11*	4	7
Respuesta construida corta	37*	18	19	28*	17	11
Respuesta construida extendida	66*	37	29	59*	30	29

* 2003 difiere significativamente ($p < .05$) de 2005.

¹ El marco teórico de NAEP de matemáticas utilizado en 2003 difiere del utilizado en 2005. La tabla presenta las áreas de contenido para 2003. En 2005, las áreas de contenido fueron numeración y operación, medición, geometría, análisis de datos y probabilidad y álgebra.

NOTA: El porcentaje promedio de respuestas faltantes se calcula utilizando las preguntas que se administraron tanto en 2003 como en 2005. Las respuestas faltantes incluyen las preguntas omitidas y las no alcanzadas. Las diferencias se basan en estimados no redondeados.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Tabla 1.4

Diferencia entre la media de las puntuaciones promedio de las preguntas en las evaluaciones NAEP de matemáticas administradas en 2003 y 2005 a estudiantes de escuelas públicas en Puerto Rico, por área de contenido y tipo de pregunta

Característica de la pregunta	4 ^o grado			8 ^o grado		
	Puntuación promedio			Puntuación promedio		
	2003	2005	Diferencia	2003	2005	Diferencia
Total¹	.27*	.28	.02	.27*	.28	.02
Área de contenido						
Numeración y operación	.27*	.29	.02	.28*	.30	.02
Medición	.28	.28	.01	.21	.22	.01
Geometría	.32*	.35	.03	.29*	.31	.02
Análisis de datos y probabilidad	.23	.23	#	.23*	.26	.02
Álgebra y funciones	.23*	.25	.02	.29*	.31	.02
Tipo de pregunta						
Selección múltiple	.33	.34	.01	.28*	.30	.01
Respuesta construida corta	.16*	.20	.04	.25*	.27	.02
Respuesta construida extendida	.03*	.07	.04	.11*	.14	.03

Se redondea a cero.

* 2003 difiere significativamente ($p < .05$) de 2005.

NOTA: La media de las puntuaciones promedio se calcula utilizando aquellas preguntas que se administraron tanto en 2003 como en 2005. Las diferencias se basan en estimados no redondeados.

¹ El marco teórico de NAEP de matemáticas utilizado en 2003 fue revisado en 2005. La tabla presenta las áreas de contenido para 2003. En 2005 las áreas de contenido fueron numeración y operación, medición, geometría, análisis de datos y probabilidad y álgebra.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.



Capítulo 2

El Desajuste de Preguntas

La evaluación NAEP de matemáticas consta de un conjunto de preguntas diseñadas para medir diferentes formas de saber y hacer matemáticas. Cada estudiante responde a un subconjunto del total de preguntas en cada una de las cinco áreas de contenido. Las respuestas se califican y el rendimiento se informa para grupos de estudiantes, no para individuos. Al utilizar modelos de la teoría de respuesta al ítem (IRT, por sus siglas en inglés), puede predecirse el rendimiento para cualquier pregunta. La relación entre el rendimiento predicho y el real es fundamental para generar la escala de NAEP para el informe de resultados. Los análisis preliminares de los resultados de 2003 en Puerto Rico mostraron una discrepancia entre el rendimiento real y el predicho para un gran número de preguntas (desajuste de preguntas). Este capítulo brinda una explicación del desajuste de preguntas y sus implicaciones en la interpretación de los resultados de Puerto Rico.

Predicción del Rendimiento de los Estudiantes

Las escalas de NAEP están diseñadas para medir la competencia con respecto a un marco teórico de conocimientos y habilidades. Las preguntas de la evaluación son ejemplos de los conocimientos y las habilidades representadas en el marco teórico. Como muchos otros programas de evaluación, NAEP hace públicas algunas preguntas y las reemplaza con nuevas en la evaluación siguiente. Por lo tanto, NAEP requiere mecanismos

para informar sobre el rendimiento en la misma escala aunque las preguntas cambien con el tiempo.

Los modelos estadísticos basados en la teoría de respuesta al ítem (IRT) se utilizan para estimar la distribución de la competencia de los estudiantes en la escala de NAEP y vinculan preguntas diferentes a una escala común (Lord 1980; Hambleton, Swaminathan y Rogers 1991; Embretson y Reise 2000). Los modelos IRT caracterizan cada pregunta de una evaluación. Para cualquier pregunta dada, la probabilidad de respuesta correcta (0.0 a 1.0) puede trazarse en un continuo de niveles de competencia.

Normalmente, la resultante curva característica de la pregunta (ICC, por sus siglas en inglés) muestra que a medida que incrementa el nivel de competencia o habilidad, incrementa también la probabilidad de una respuesta correcta.² Los estudiantes más competentes (mayor habilidad en matemáticas) son más propensos a responder correctamente a una pregunta que los estudiantes menos competentes.

La forma de la ICC está determinada por la dificultad y la discriminación de la pregunta. Cuanto más difícil una pregunta, más baja la probabilidad de que respondan correctamente estudiantes con menor conocimiento y destrezas en matemáticas (menor habilidad o competencia). La discriminación de la pregunta se refleja en lo empinado de la ICC. Así, cuanto más empinada la curva (pendiente), mejor puede la pregunta discriminar entre aquellos con más o menos competencia. Estas características de la ICC son útiles para examinar la concordancia entre el rendimiento esperado y el rendimiento real en un conjunto de preguntas.

Desajuste de Preguntas

Una ICC establece el rendimiento esperado de los estudiantes. El ajuste de pregunta se refiere hasta qué punto realmente los estudiantes se desempeñan como se espera en una pregunta. El ajuste de pregunta se evalúa comparando la proporción de estudiantes examinados, dentro de un rango relativamente estrecho de competencia que responde correctamente a la pregunta con el rendimiento esperado en ese rango. La discrepancia entre el rendimiento esperado y el rendimiento real se define como desajuste de pregunta.

En Puerto Rico, algunas preguntas reflejaron desajuste. Para ilustrar el concepto de desajuste de pregunta es útil primero considerar la ICC para una pregunta con ajuste perfecto (Figura 2.1). En la figura, cada círculo representa el rendimiento real de un conjunto de estudiantes examinados. Los que se encuentran en el extremo inferior de la escala de competencia tienen una baja probabilidad de respuesta correcta a la pregunta.

² Theta se utiliza para denotar habilidad o competencia en el constructo subyacente que la evaluación intenta medir / estimar. La medida de habilidad no tiene unidades de medición inherentemente significativas. Lo más común, aunque arbitrario, es usar la escala theta en la que cero es la media y 1 es la desviación estándar, la escala de puntuación z.

A medida que la curva se desplaza hacia la derecha, los estudiantes con mayor competencia tienen una mayor probabilidad de responder correctamente a esta pregunta. En esta figura, el rendimiento real (círculos) cae exactamente sobre la ICC.

La Figura 2.2 muestra una discrepancia entre el rendimiento real y el esperado. Los círculos encima de la línea muestran grupos de estudiantes cuyo rendimiento fue mejor de lo esperado, y los círculos que están debajo de la línea muestran estudiantes cuyo rendimiento fue inferior a lo esperado para esta pregunta. Como se indicó arriba, para una pregunta con ajuste perfecto, se espera que la mayoría de los estudiantes menos competentes contesten incorrectamente la pregunta y se espera que la mayoría de los estudiantes más competentes la contesten correctamente. En la Figura 2.2, la probabilidad de una respuesta correcta fue mayor o menor de lo esperado para un nivel de competencia dado.

Figura 2.1
Ejemplo de la curva característica de una pregunta con ajuste perfecto

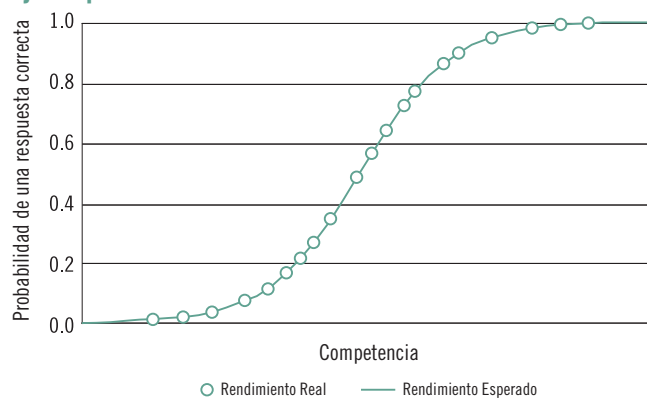
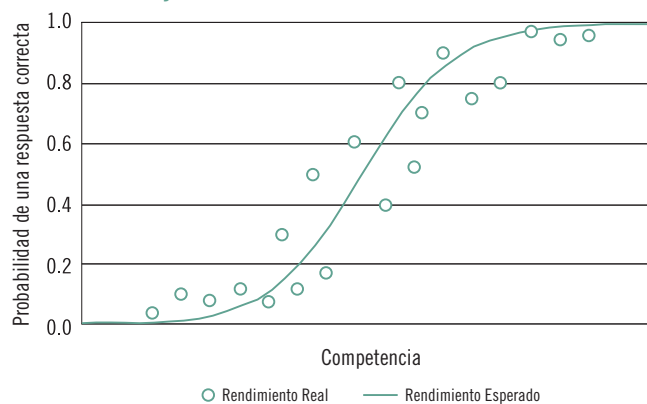


Figura 2.2
Ejemplo de la curva característica de una pregunta que muestra desajuste



Fuentes de Desajuste de Preguntas

El rendimiento esperado para las preguntas de NAEP de matemáticas de 2003 se estimó a partir de la muestra de Estados Unidos (todas las jurisdicciones participantes excepto Puerto Rico). Los niveles elevados de desajuste de preguntas en Puerto Rico pueden indicar diferencias curriculares entre Puerto Rico y Estados Unidos, o errores de traducción, o una evaluación muy difícil, o alguna combinación de las anteriores.

- **Diferencias curriculares.** Las jurisdicciones difieren en cuándo enseñan ciertas habilidades o conceptos. Por ejemplo, una jurisdicción puede enseñar multiplicación de fracciones en cuarto grado, mientras que otra las enseña en quinto grado. Un estudiante competente en matemáticas, pero al que aún no se le ha enseñado multiplicación de fracciones, tendría dificultad resolviendo una pregunta de este tipo.

- **Traducción.** En Puerto Rico, NAEP se administra utilizando un instrumento en español. Para la evaluación de 2003, el proceso de traducción estuvo centrado en desarrollar una sola versión al español que pudiera administrarse universalmente a estudiantes de ascendencia puertorriqueña, cubana, mexicana, centroamericana y española. En evaluaciones subsiguientes, un proceso de revisión de la traducción

evaluó si las traducciones eran adecuadas para administrarse únicamente en Puerto Rico. El proceso de traducción de la evaluación, del inglés al español, puede resultar en cambios sutiles que alteran el significado de una pregunta, o hacer que las preguntas sean más fáciles o más difíciles.

- **Incompatibilidad entre la dificultad de la evaluación y la competencia de los estudiantes.**

La dificultad de una pregunta se estimó a base de preguntas de prueba con la muestra de Estados Unidos. Antes de la administración de 2003 en Puerto Rico las preguntas no se pusieron a prueba. Algunas de las preguntas podrían haber tenido un nivel de dificultad diferente para los estudiantes de Puerto Rico que para los estudiantes de los Estados Unidos. Una batería de preguntas de la evaluación que es más difícil que los niveles de competencia de los estudiantes puede incrementar las respuestas no contestadas, las respuestas adivinadas y trastocar competencia con velocidad. El análisis preliminar de los datos de 2003 mostró niveles más elevados de respuestas faltantes y más respuestas incorrectas para los estudiantes en Puerto Rico comparado con los estudiantes de la muestra de Estados Unidos.

Implicaciones del Desajuste de Preguntas

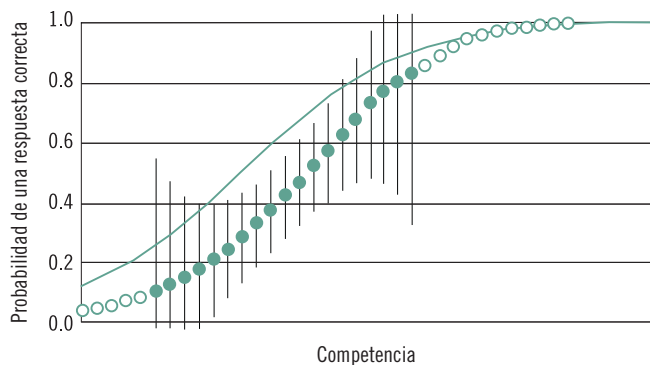
Aunque muchas preguntas de la evaluación NAEP no funcionaron como se esperaba en Puerto Rico, esto no impide que se informen los datos de Puerto Rico en la escala de NAEP. El desajuste de preguntas apunta a una discrepancia entre el rendimiento real y el predicho, y sugiere la necesidad de nuevos análisis para (1) entender la naturaleza del desajuste de pregunta y (2) evaluar las implicaciones del desajuste de pregunta para informar los resultados de Puerto Rico en la escala de NAEP.

Estos asuntos pueden abordarse examinando las ICC de preguntas que no funcionan como se espera. Las gráficas de ejemplo 2.3 y 2.4 incluyen barras de error estándar que muestran el margen de error alrededor de cada punto. Cuando el margen de error es demasiado grande para visualizarse, los puntos se muestran como círculos vacíos, lo cual indica que no hay suficientes estudiantes en ese rango de competencia para proporcionar datos válidos.



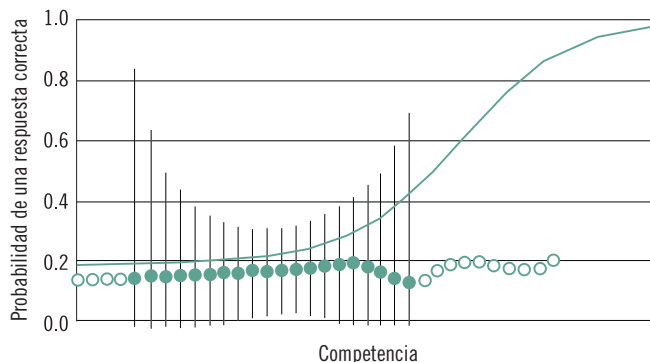
• **Preguntas demasiado difíciles para la población.** La Figura 2.3 muestra un ejemplo de pregunta que es demasiado difícil para los estudiantes examinados. Hay dos aspectos importantes a notar aquí. Primero, el rendimiento real de los estudiantes (círculos) está por debajo de la curva del rendimiento predicho, salvo en los niveles más elevados de competencia. Segundo, muy pocos estudiantes en Puerto Rico están en el extremo más elevado de la escala de competencia. Preguntas como éstas no contribuyen mucho a los estimados de competencia general de los estudiantes porque la gran mayoría de ellos las contesta incorrectamente.

Figura 2.3
Ejemplo de curva característica de una pregunta que es muy difícil para los estudiantes



NOTA: Los círculos indican el rendimiento real de los estudiantes. Los círculos vacíos indican un número insuficiente de estudiantes en un nivel de competencia para proporcionar datos válidos.

Figura 2.4
Ejemplo de curva característica de una pregunta relativamente no relacionada a la competencia de los estudiantes



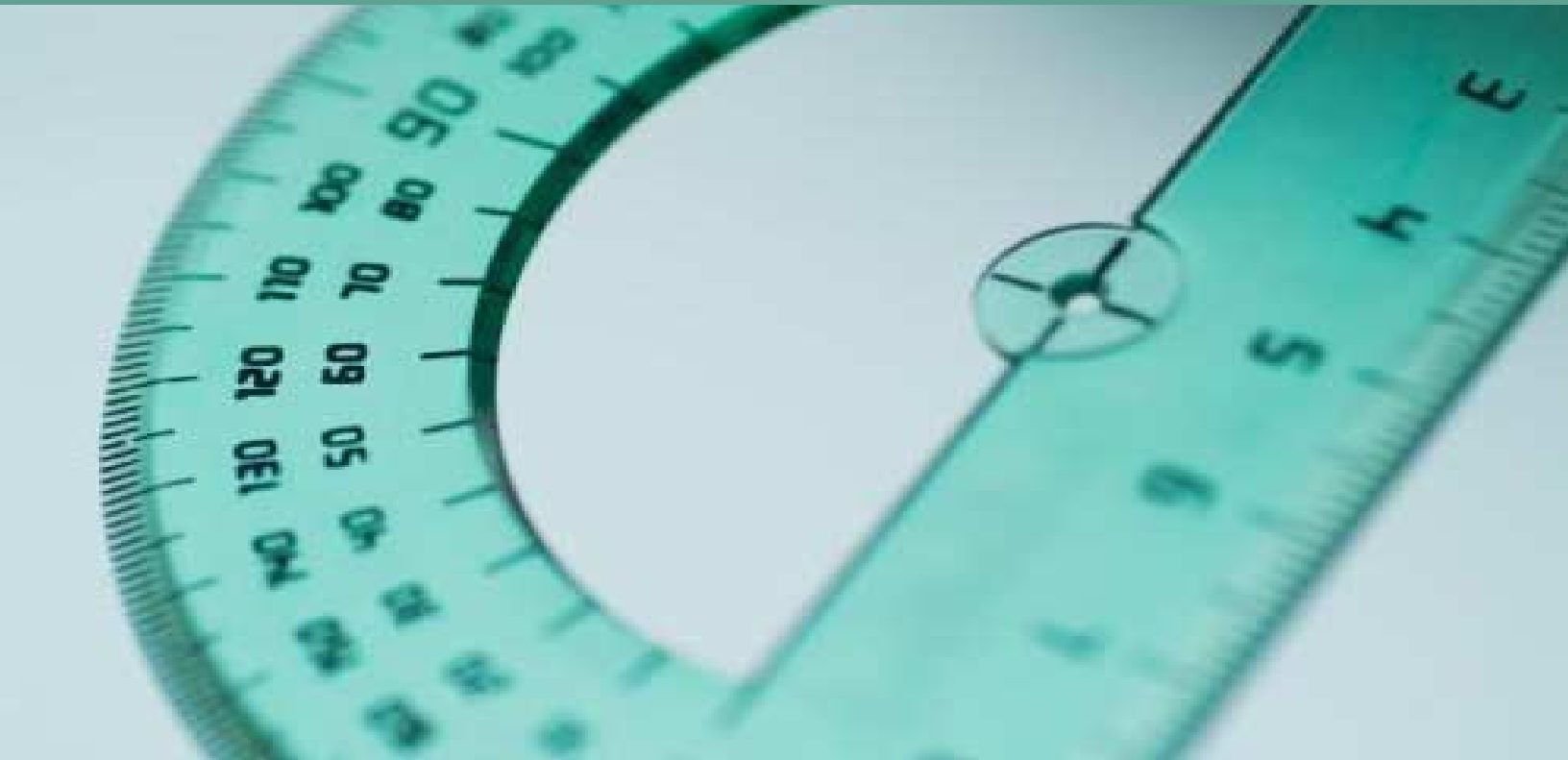
NOTA: Los círculos indican el rendimiento real de los estudiantes. Los círculos vacíos indican un número insuficiente de estudiantes en un nivel de competencia para proporcionar datos válidos.



• **Rendimiento en la pregunta no relacionado a competencia.** La Figura 2.4 presenta una pregunta relativamente no relacionada a la competencia de los estudiantes. Se espera que a mayor rendimiento en la pregunta mayor competencia (curva ascendente a la derecha de la gráfica). Sin embargo, en Puerto Rico el rendimiento real de los estudiantes (indicado por los círculos) es relativamente plano. Para esta pregunta, la probabilidad de responder correctamente no está relacionada a la competencia de los estudiantes. Los estudiantes con más conocimiento y habilidad en matemáticas (extremo superior de la escala de competencia) tienen igual probabilidad de responder correctamente a esta pregunta que los estudiantes con menos conocimiento y habilidad (extremo inferior de la escala de competencia).

La pregunta puede haber sido difícil para los estudiantes en Puerto Rico porque el concepto no había sido suficientemente presentado antes de la evaluación. Por otra parte, la pregunta puede haber sido difícil de entender para los estudiantes en Puerto Rico debido a errores de traducción del inglés al español que alteraron el significado de la pregunta, o quizás la traducción introdujo más de una respuesta correcta o casi correcta. Tales preguntas no crean sesgos ya que no conducen a puntuaciones sistemáticamente mayores o menores, pero sí aumentan el error de medición (precisión con la que se puede estimar el rendimiento).

En resumen, muchas preguntas de NAEP funcionaron distinto a lo esperado en Puerto Rico. Al reducir la precisión de los estimados, el desajuste de pregunta afecta principalmente las estadísticas agregadas. El desajuste de pregunta puede indicar problemas con la evaluación y resultados potencialmente sesgados. Sin embargo, los resultados pueden ser un reflejo preciso de la competencia del grupo que se evalúa. El capítulo 3 examina la precisión de los resultados de Puerto Rico en términos de la escala de NAEP.



Capítulo 3

Informe de los Resultados de Puerto Rico en la Escala de NAEP

Los resultados de Puerto Rico para 2003 y 2005 no se informaron con los de otras jurisdicciones debido a las inquietudes de que las preguntas traducidas no estuvieran funcionando como se esperaba. Se realizó un conjunto de análisis para examinar si la evaluación NAEP de matemáticas medía el logro de los estudiantes en Puerto Rico del mismo modo en que medía el logro de los estudiantes en Estados Unidos. Este capítulo describe los análisis y las implicaciones para informar los resultados de Puerto Rico en la escala de NAEP.

En futuras administraciones de NAEP la intención es incluir a Puerto Rico como parte de la muestra de Estados Unidos. Aunque Puerto Rico participó en las evaluaciones NAEP de matemáticas de 2003 y 2005, los resultados no se incluyeron en los estimados generales de Estados Unidos debido a inquietudes sobre la calidad de los datos. En Puerto Rico hubo una mayor cantidad de datos faltantes, menos respuestas correctas y niveles más elevados de desajuste de preguntas que en Estados Unidos. La publicación de los resultados de 2003 y 2005 se demoró en espera de análisis adicionales para examinar la precisión de los resultados de Puerto Rico en términos de la escala de NAEP. Los

análisis se realizaron en cuatro pasos. Primero, las preguntas identificadas como problemáticas para Puerto Rico en la evaluación NAEP de matemáticas de 2003 y 2005 se eliminaron del conjunto total de preguntas administradas para producir un conjunto reducido de preguntas. Segundo, la cobertura de contenido en términos de las cinco áreas de contenido de NAEP de matemáticas se comparó para el conjunto total de preguntas y para el conjunto reducido de preguntas. Tercero, para permitir comparaciones los dos conjuntos de preguntas fueron calibrados y ubicados en la escala de NAEP. Cuarto, el rendimiento en el conjunto total de preguntas se comparó con el rendimiento en el con-

junto reducido de preguntas para Puerto Rico, Estados Unidos y las jurisdicciones que componen la muestra de Estados Unidos. Al llevar a cabo estos análisis, se establecen comparaciones para estudiantes de escuelas públicas ya que sólo estudiantes de escuelas públicas participaron en Puerto Rico en la evaluación NAEP de matemáticas de 2003 y 2005.

Identificación de Preguntas Problemáticas

Las preguntas se identificaron como problemáticas por una de dos razones: (1) errores de traducción y funcionamiento diferencial de las preguntas, o (2) preguntas que no funcionaron bien en Puerto Rico.

Errores de traducción y funcionamiento diferencial de las preguntas. Una posible explicación del desajuste de preguntas son errores de traducción. Tras la preparación de los folletos de prueba de 2003, a solicitud de NCES, una compañía de traducciones realizó una revisión independiente de la traducción. El proceso de verificación de la traducción clasificó las preguntas como sin errores de traducción o con errores menores, errores moderados o errores graves. Sin embargo, ni siquiera errores de traducción graves justificaron la mayoría del desajuste de preguntas en Puerto Rico.

Todas las preguntas utilizadas en la evaluación NAEP de matemáticas de 2003 pasaron por un análisis de funcionamiento diferencial de la pregunta (DIF, por sus siglas en inglés),³ una técnica para identificar el desajuste de pregunta. Para cada análisis DIF, el rendimiento en Puerto Rico (el grupo focal) se comparó con el rendimiento de los siguientes grupos de referencia: Estados Unidos, Distrito de Columbia, Islas Vírgenes, Samoa norteamericana y estudiantes de la muestra pública de Estados Unidos que se identificaron a sí mismos como puertorriqueños. El objetivo fue determinar si las preguntas funcionaban de modo distinto en Puerto Rico que en Estados Unidos u otras jurisdicciones por razones no vinculadas a la habilidad.

³ Una pregunta presenta DIF si la probabilidad de responderla bien depende de la pertenencia a un grupo, incluso tras regular la diferencia de habilidad. Los métodos DIF utilizados en este análisis fueron el procedimiento de chi-cuadrado de Mantel-Haenszel para preguntas dicotómicas y el procedimiento Mantel para preguntas politómicas (Holland y Wainer 1993).



Los resultados indicaron que las preguntas señaladas con un DIF significativo no justificaron ni el total ni la mayoría del desajuste de preguntas.

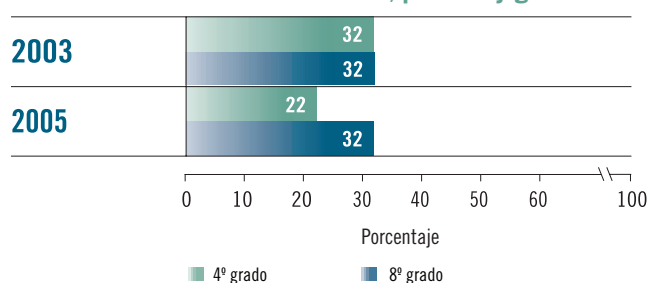
Las preguntas de la evaluación NAEP de matemáticas de 2003 identificadas con (1) graves errores de traducción y (2) un DIF significativo se consideraron problemáticas. Los errores de traducción identificados en 2003 fueron corregidos antes de la administración de NAEP de 2005 en Puerto Rico. Consecuentemente, la revisión de la traducción y el análisis DIF de las preguntas con errores de traducción no se repitieron para los datos de 2005.

Preguntas que no funcionaron bien en Puerto Rico. En 2003 y 2005, la mayoría de las preguntas identificadas como problemáticas no funcionaron bien para la muestra de Puerto Rico. Éstas incluyeron (1) preguntas que no estaban correlacionadas con la puntuación total para la muestra de Puerto Rico; (2) preguntas sin varianza en respuesta para la muestra de Puerto Rico; (3) preguntas que no discriminaban entre estudiantes con mayor o menor habilidad (ICC plano para Puerto Rico); y (4) preguntas que mostraron desajuste entre las ICC teóricas y empíricas para la muestra de Puerto Rico.

Cobertura del Contenido del Conjunto Total y el Conjunto Reducido de Preguntas

Las preguntas problemáticas se eliminaron del conjunto total de preguntas para crear un conjunto reducido de preguntas para cuarto y octavo grado en 2003 y 2005. En promedio, entre años y grados, 30 por ciento de las preguntas se eliminó del conjunto total de preguntas (Figura 3.1).

Figura 3.1
Porcentaje del total de preguntas eliminadas de la evaluación NAEP de matemáticas, por año y grado



FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Tabla 3.1
Distribución porcentual del conjunto total y de los conjuntos de preguntas eliminadas y conservadas en 4º grado para la evaluación NAEP de matemáticas en 2003, por característica de la pregunta

Característica de la pregunta	Eliminada	Conservada	Completa
Área de contenido			
Numeración y operación	32	47	42
Medición	19	17	18
Geometría	18	14	15
Análisis de datos y probabilidad	11	11	11
Álgebra y funciones	21	11	15
Tipo de pregunta*			
Selección múltiple	79	57	64
Respuesta construida corta	18	39	32
Respuesta construida extendida	4	5	4
Habilidad en matemáticas			
Comprensión de conceptos	46	37	40
Resolución de problemas	40	34	36
Conocimiento de procedimientos	14	29	24

* La distribución de las preguntas eliminadas difiere significativamente ($p < 0.05$) de la distribución de las preguntas conservadas.

NOTA: Los datos desglosados pueden no corresponder al total debido al redondeo.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003.

La eliminación de preguntas del conjunto total de preguntas podría alterar el constructo que se está midiendo o podría desplazar el énfasis de un área de contenido a otra. Debido a que el conjunto reducido de preguntas es un subconjunto del conjunto total de preguntas, se utilizaron análisis de chi-cuadrado para comparar la distribución de las preguntas conservadas (preguntas que componen el conjunto reducido) con la distribución de las preguntas eliminadas (total de preguntas menos preguntas conservadas). Como se muestra en la Tabla 3.1, la distribución porcentual de las preguntas por área de contenido y por habilidad matemática para el conjunto reducido de preguntas no difiere significativamente de la correspondiente distribución de preguntas para el conjunto eliminado de preguntas utilizado en la administración de NAEP de 2003. El porcentaje de preguntas por tipo de pregunta en el subconjunto reducido difirió significativamente de la distribución por tipo de pregunta en el conjunto eliminado. Para 2005 el patrón de resultados fue el mismo en cuarto grado. En octavo grado la distribución de preguntas eliminadas por área de contenido, tipo de pregunta o habilidad matemática no difirió

significativamente de la distribución de preguntas conservadas en 2003 ó 2005. El Apéndice A incluye información adicional sobre el conjunto completo y los conjuntos de preguntas eliminadas y de preguntas conservadas.

La Tabla A-1 muestra la distribución porcentual del conjunto total y de los conjuntos de preguntas eliminadas y conservadas para cuarto grado, por año y características de las preguntas. La Tabla A-2 muestra la distribución porcentual del conjunto total y de los conjuntos de preguntas eliminadas y conservadas para octavo grado, por año y características de las preguntas. La Tabla A-3 muestra las puntuaciones promedio mínimas y máximas del conjunto total de preguntas y del conjunto reducido para Puerto Rico y Estados Unidos, por grado y año.

Calibración de la Escala Completa y de la Escala Restringida

Se crearon dos escalas—una escala completa y una escala restringida. La escala completa se basó en el conjunto total de preguntas de NAEP de matemáticas administradas en 2003 y 2005. La escala restringida se basó en un conjunto reducido de preguntas (aquellas preguntas que quedaron después de excluir las preguntas problemáticas). Para la escala restringida fue necesario recalibrar la evaluación para contemplar los cambios en el tratamiento de las preguntas (exclusión de preguntas problemáticas) y la inclusión de los estudiantes de Puerto Rico. De conformidad con los procedimientos utilizados en NAEP operacional, las cinco subescalas de matemáticas se calibraron por separado y luego se combinaron en una combinación lineal.

Las constantes de enlace se calcularon para transformar los resultados de la escala restringida de 2003 a los de la escala completa en 2003, ubicándolos así en la escala de NAEP y permitiendo la comparación de resultados.⁴ Los resultados de la escala restringida de 2005 se enlazaron luego a los resultados de la escala restringida de 2003 utilizando procedimientos estándar de NAEP. La metodología utilizada para equiparar las puntuaciones durante este procedimiento de enlace puede introducir error

⁴NAEP utiliza procedimientos comunes de enlace de poblaciones (Allen, Donoghue y Schoeps 2001).

en los estimados de la habilidad de los estudiantes. Sin embargo, para NAEP no se han desarrollado procedimientos estadísticos y metodología necesaria para estimar adecuadamente el impacto del error de enlace en este contexto.

Comparación de la Escala Completa y la Escala Restringida

Se llevaron a cabo comparaciones de la escala completa y la escala restringida para Puerto Rico, Estados Unidos y jurisdicciones individuales que componen la muestra de Estados Unidos. Tres preguntas guiaron este análisis. Primero: ¿Hay alguna diferencia en la puntuación promedio general? Segundo: ¿Varían por escala los porcentajes de estudiantes con rendimiento en cada nivel de logro? Las diferencias en puntuaciones promedio generales pueden o no afectar los porcentajes de estudiantes con rendimiento en cada nivel de logro. Tercero: ¿Son consistentes los resultados de las comparaciones entre la escala completa y la restringida entre jurisdicciones con los resultados de estas comparaciones en Puerto Rico y los Estados Unidos? Considerados juntos, los resultados de estos análisis aportan evidencia que sustenta el informe de los resultados de Puerto Rico en la escala de NAEP, permitiendo así comparaciones entre Puerto Rico y Estados Unidos.





Puntuaciones promedio en escala. Utilizando la escala completa y la escala restringida se compararon las puntuaciones promedio de NAEP de matemáticas para Puerto Rico y para Estados Unidos (Tabla 3.2).⁵ Para Puerto Rico, la diferencia entre la puntuación promedio en la escala completa y la escala restringida varía entre 1.0 y 3.4 puntos entre grados y administraciones de NAEP. En Puerto Rico, las puntuaciones promedio fueron más elevadas en la escala restringida que en la escala completa. Para Estados Unidos, la puntuación promedio en las dos escalas concuerda dentro de las tres décimas de un punto de escala en cuarto y octavo grado, tanto para 2003 como para 2005. La diferencia entre las puntuaciones promedio en la escala completa y la escala restringida fue estadísticamente significativa en 2005. La significación estadística probablemente se debe al gran tamaño de las muestras.

⁵ Las diferencias entre puntuaciones en escala o porcentajes se calcularon utilizando números no redondeados. En algunas instancias, el resultado de la resta difiere de lo que se obtendría restando los valores redondeados que se muestran en la figura o tabla correspondiente.

Tabla 3.2

Puntuaciones promedio de la escala completa y la escala restringida para la evaluación NAEP de matemáticas en Estados Unidos y en Puerto Rico, por año y grado

Jurisdicción	Escala completa		Escala restringida		Escala completa menos escala restringida	
	Promedio	Error estándar	Promedio	Error estándar	Promedio	Error estándar
Puerto Rico						
2003						
4º grado	179	1.0	182	1.0	-3.0*	0.13
8º grado	212	1.0	214	1.1	-2.0*	0.16
2005						
4º grado	183	0.9	187	0.9	-3.4*	0.13
8º grado	218	1.0	220	1.1	-1.9*	0.11
EE. UU.						
2003 ¹						
4º grado	234	0.2	234	0.2	0.0	0.02
8º grado	276	0.3	276	0.3	0.0	0.04
2005						
4º grado	237	0.2	237	0.2	0.3*	0.02
8º grado	278	0.2	277	0.2	0.1*	0.03

* Diferencia estadísticamente significativa ($p < .05$).

¹ La concordancia perfecta entre la escala completa y la restringida en 2003 es producto del procedimiento de enlace.

NOTA: Los errores estándar no incluyen el error de enlace, por lo que pueden subestimar los verdaderos errores estándar. Para NAEP no se han desarrollado procedimientos estadísticos y metodología necesaria para estimar adecuadamente el impacto del error de enlace.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Niveles de logro. NAEP informa el rendimiento a base de tres niveles de logro: *Básico*, *Competente* y *Avanzado*. Como pocos estudiantes en Puerto Rico se desempeñaron en el nivel *Avanzado*, sólo se proporcionan datos para los niveles *Básico* y *Competente*. Los detalles están disponibles en *Matemáticas 2003 y 2005, Rendimiento en Puerto Rico: Aspectos Sobresalientes*, disponible en http://nationsreportcard.gov/puertorico_2005/.

La Tabla 3.3 presenta el porcentaje de estudiantes en o por encima del nivel *Básico* y el porcentaje de estudiantes en o por encima del nivel *Competente*. En Puerto Rico, un mayor porcentaje de estudiantes recibió una puntuación en o por encima del nivel *Básico* en la escala restringida que en la escala completa. Las diferencias variaron entre 1.16 y 1.71 puntos porcentuales en cuarto y octavo grado en 2003 y 2005. En Estados Unidos, menos estudiantes recibieron una puntuación en o por encima del nivel *Básico* en la escala restringida que en la escala completa. Las diferencias variaron entre 0.32 y 0.82 puntos porcentuales por grado en 2003 y 2005.



Tabla 3.3

Porcentaje de estudiantes que se desempeñaron en seleccionados niveles de logro de NAEP en la escala completa y la restringida, en Estados Unidos y en Puerto Rico, por grado y año

Jurisdicción	4º grado				8º grado			
	2003		2005		2003		2005	
	Completa	Restringida	Completa	Restringida	Completa	Restringida	Completa	Restringida
Puerto Rico								
Porcentaje por debajo de <i>Básico</i>	91*	89	88	87	96*	94	94*	93
Porcentaje en o por encima de <i>Básico</i>	9*	11	12*	13	4*	6	6*	7
Porcentaje en o por encima de <i>Competente</i>	#	#	#	#	#	#	#	#
EE. UU.								
Porcentaje por debajo de <i>Básico</i>	24	24	21	21	33	33	32	32
Porcentaje en o por encima de <i>Básico</i>	76	76	79	79	67	67	68	68
Porcentaje en o por encima de <i>Competente</i>	31*	31	35*	35	27*	27	28*	28

Se redondea a cero.

* La escala completa difiere significativamente ($p < .05$) de la escala restringida.

NOTA: Las diferencias entre porcentajes se calculan utilizando números no redondeados. En algunas instancias, el resultado de la resta difiere de lo que se obtendría restando los números redondeados que se muestran en la tabla.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.



Comparaciones entre otras jurisdicciones. El rendimiento de las jurisdicciones individuales que componen la muestra de Estados Unidos se resume en la Tabla 3.4. La hilera 1 muestra la diferencia promedio en general entre la escala completa y la restringida para la muestra de Estados Unidos. Entre las jurisdicciones que componen la muestra de Estados Unidos, las diferencias máximas y mínimas de las puntuaciones promedio en escala se muestran en las hileras 2 y 3, respectivamente. La hilera 4 muestra el porcentaje de jurisdicciones participantes con diferencias de puntuaciones promedio inferiores a un punto de escala.

Como se muestra en la Tabla 3.4, las diferencias entre puntuaciones promedio en la escala completa y la restringida varían entre 0.5 y 1.6 puntos entre jurisdicciones que componen la muestra de Estados Unidos. Las diferencias máximas y mínimas, mostradas en las hileras 2 y 3 respectivamente, indican que algunos estados tuvieron puntuaciones promedio más elevadas en la escala completa y otros estados tuvieron puntuaciones promedio más elevadas en la escala restringida. Por ejemplo, en cuarto grado en 2003, una jurisdicción tuvo una puntuación promedio en la escala completa 1.2 puntos por debajo de su puntuación promedio en la escala restringida, y otra jurisdicción tuvo una puntuación promedio en la escala completa 0.5 puntos por encima de su puntuación promedio en la escala restringida. En comparación, las puntuaciones promedio en la escala completa para Puerto Rico variaron entre 1.9 y 3.4 puntos por debajo de las puntuaciones promedio en la escala restringida (Tabla 3.2).

La cuarta hilera de la Tabla 3.4 muestra el porcentaje de jurisdicciones para las cuales los promedios en la escala completa y en la escala restringida difieren en menos de 1 punto de escala. Los porcentajes elevados indican que se pueden comparar las puntuaciones promedio utilizando las dos escalas. Además de puntuaciones promedio comparables en la escala completa y la restringida, las correlaciones entre las puntuaciones de escala en cuarto y octavo grado en 2003 y 2005 fueron elevadas (promedio 0.99).

Tabla 3.4

Indicadores de concordancia entre la escala completa y la restringida para NAEP de matemáticas, por año y grado

Indicador	2003		2005	
	4º grado	8º grado	4º grado	8º grado
Diferencia entre el estimado en la escala completa y la restringida para la muestra de Estados Unidos	#	#	0.3*	0.1*
Diferencia máxima entre las puntuaciones promedio para la escala completa y la restringida entre las jurisdicciones	0.5	0.5	0.9	1.6
Diferencia mínima entre las puntuaciones promedio para la escala completa y la restringida entre las jurisdicciones	-1.2	-1.1	-0.7	-1.0
Porcentaje de jurisdicciones con diferencias de la escala completa menos la restringida inferiores a 1 punto	98.0	98.0	100.0	94.0

Se redondea a cero.

* La diferencia entre estimados en la escala completa y la restringida es estadísticamente significativa ($p < .05$).

NOTA: Los errores estándar no incluyen el error de enlace, por lo que pueden subestimar los verdaderos errores estándar. Para NAEP no se han desarrollado procedimientos estadísticos y metodología necesaria para estimar adecuadamente el impacto del error de enlace.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Conclusiones

Se realizó un conjunto de análisis para examinar si la evaluación NAEP de matemáticas medía el logro de los estudiantes en Puerto Rico del mismo modo que medía el logro de los estudiantes en Estados Unidos. ¿Pueden los resultados de la evaluación NAEP de matemáticas de 2003 y 2005 en Puerto Rico informarse en la escala de NAEP? Para responder a este interrogante las preguntas identificadas como problemáticas en Puerto Rico se eliminaron del conjunto total de preguntas. Se compararon los dos conjuntos de preguntas, el conjunto eliminado de preguntas de NAEP administradas y el conjunto conservado de preguntas (aquellas que quedaron después de eliminarse las problemáticas). Los resultados de la comparación indicaron que los dos conjuntos de preguntas son similares en términos de la distribución de preguntas por contenido y habilidad en matemáticas en cuarto y octavo grado, tanto en 2003 como en 2005. La distribución de preguntas por tipo de pregunta en cuarto grado (pero no en octavo grado) en el conjunto eliminado difirió estadísticamente de la distribución de preguntas por tipo de pregunta en el conjunto conservado en 2003 y 2005. Las preguntas fueron calibradas y las escalas fueron enlazadas utilizando procedimientos estándar de NAEP, permitiendo así comparaciones entre escalas.

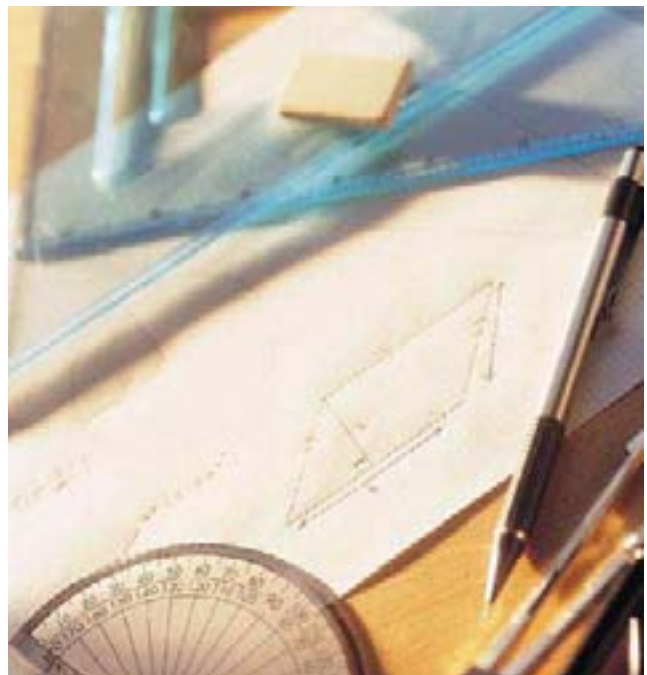
En Estados Unidos, las comparaciones de la escala completa y la escala restringida indican: (1) las puntuaciones promedio en escala concuerdan a tres décimas de un punto de escala, (2) el porcentaje de estudiantes que se desempeñan en diversos niveles de logro concuerda a 1 punto de porcentaje y (3) la correlación entre puntuaciones promedio en escala para jurisdicciones que componen la muestra de Estados Unidos es .99 en cuarto y octavo grado en 2003 y 2005. Estos resultados indican que se pueden comparar la escala completa y la escala restringida para Estados Unidos y las jurisdicciones que componen la muestra de Estados Unidos.

Para Puerto Rico, la diferencia entre las puntuaciones promedio en las dos escalas es de 2 a 3 puntos de escala, y la diferencia entre el porcentaje de estudiantes

que alcanza un nivel de logro en particular en las dos escalas es de 1 a 2 puntos porcentuales. Las puntuaciones promedio de Puerto Rico fueron más elevadas en la escala restringida que en la escala completa.

¿Qué significa esto para presentar e interpretar los resultados de las administraciones de NAEP de 2003 y 2005 en Puerto Rico? Basándose en estos análisis, la conclusión de NCES es que el conjunto total de preguntas de NAEP puede y debe utilizarse para informar sobre los resultados de Puerto Rico, tanto para la evaluación de 2003 como para la de 2005. La puntuación promedio en escala para Puerto Rico podría ser entre 2 y 3 puntos mayor que la informada, pero esta diferencia no alteraría el *ranking* de Puerto Rico en relación con otras jurisdicciones participantes. Asimismo, para los resultados en niveles de logro, los porcentajes de estudiantes en o por encima de un nivel de logro en particular pueden discrepar para las dos escalas en 1 ó 2 puntos porcentuales.

No sería adecuado utilizar la escala restringida para Estados Unidos y Puerto Rico porque podrían medirse menos objetivos del marco teórico utilizando la escala restringida. La cobertura de contenido y la estabilidad de los estimados a lo largo del tiempo son consideraciones clave al encaminarse Puerto Rico hacia la integración completa al programa NAEP.





Capítulo 4

Integración de Puerto Rico a NAEP: Próximos Pasos

En futuras administraciones de NAEP el objetivo es incluir a Puerto Rico en la muestra de Estados Unidos e informar los resultados de Puerto Rico junto con los de otras jurisdicciones participantes. La administración de 2007 en Puerto Rico siguió los procedimientos utilizados en la administración de 2005. Se han de implementar planes para incorporar a Puerto Rico al muestreo, la recopilación de datos y el informe para la administración de 2009 de NAEP. El acceso oportuno a los resultados de NAEP puede contribuir a los esfuerzos de padres, educadores y legisladores para mejorar los logros de los estudiantes.

Administración de 2007

La evaluación NAEP de matemáticas de 2007 se llevó a cabo entre enero y marzo. Salvo dos excepciones, los procedimientos de muestreo y recopilación de datos fueron consistentes con los de otras jurisdicciones que participaron en NAEP en 2007. Primero, en Puerto Rico toda la administración se realizó en español. Segundo, a los estudiantes se les concedieron 10 minutos adicionales para completar cada sección

de la evaluación. Estos mismos procedimientos de administración se utilizaron en Puerto Rico en 2005.

Además de la administración de 2007, NAEP llevó a cabo una prueba piloto de nuevas preguntas que, por primera vez, incluyó a una muestra de estudiantes en Puerto Rico. Los resultados de la prueba piloto se utilizarán junto con otra información para hacer la selección final de preguntas para la evaluación de 2009.

Próximos Pasos

En preparación para la evaluación NAEP de matemáticas de 2009, se están realizando esfuerzos para incrementar la participación de educadores de Puerto Rico en el desarrollo y la traducción de las preguntas de matemáticas. Estos son importantes pasos siguientes hacia el objetivo de integrar plenamente a Puerto Rico en el programa NAEP.

• *Incrementar la participación de representantes de Puerto Rico en el desarrollo de evaluaciones.*

En preparación para una evaluación operacional, el comité de desarrollo de NAEP se reúne para revisar las preguntas antes de la prueba piloto, asesorar acerca del desarrollo de guías de evaluación y examinar los resultados de pruebas piloto. El comité de desarrollo de NAEP incluirá a un representante de Puerto Rico que esté familiarizado con los aspectos lingüísticos, culturales y curriculares de la enseñanza de matemáticas que puedan ser particulares de Puerto Rico. Aunque el marco teórico de NAEP continúa siendo la base sobre la que se desarrollan las preguntas de la evaluación, el representante de Puerto Rico podría señalar preguntas con contexto desconocido o uso del lenguaje que podría interferir con la medición precisa de los conocimientos y las habilidades de los estudiantes.

Segundo, el programa NAEP incluirá a representantes del Departamento de Educación de Puerto Rico en las revisiones por parte de los estados de las preguntas que tienen lugar antes de cada prueba piloto. Desde la primera evaluación NAEP en 1990, estas revisiones de preguntas por parte de los estados se han transformado en un chequeo importante de posibles cuestiones de equidad relacionadas con las variaciones entre estados y regiones. Los comentarios e inquietudes de los representantes durante estas revisiones de las preguntas pueden dar lugar a la modificación de preguntas o a la exclusión de preguntas de las pruebas piloto.

• *Incluir más educadores de Puerto Rico en el proceso de traducción.* Los procedimientos de revisión y verificación de traducción para la administración de NAEP de 2005 incluyeron a dos maestros de Puerto Rico, uno de cuarto grado y uno de matemáticas de octavo grado. Dado que se desarrolla una versión en español de la evaluación NAEP de matemáticas para uso exclusivo en Puerto Rico, algunos de los expertos que participen en los procedimientos de traducción y verificación serán nativos de Puerto Rico e involucrados actualmente en la enseñanza de matemáticas a estudiantes de la isla. Es altamente deseable continuar mejorando la comprensión de asuntos generales relacionados con la traducción y adaptación de un lenguaje a otro manteniendo el constructo original a evaluarse.

Se han implementado planes para incorporar a Puerto Rico al muestreo, la recopilación de datos y el informe para la administración de NAEP de 2009. El incrementar la participación de Puerto Rico en diversas etapas del desarrollo y revisión de la evaluación ayudará a desarrollar una evaluación que pueda adaptarse para usarse en Puerto Rico sin alterar el constructo que se evalúa.



Capítulo 5

Muestreo e Implementación

Procedimientos de Muestreo de NAEP en Puerto Rico

Las escuelas y estudiantes que participan en la evaluación NAEP son seleccionados para que sean nacionalmente representativos. El muestreo se realizó en dos etapas. En la primera etapa, dentro de cada jurisdicción se seleccionaron escuelas a partir de marcos estratificados. En la segunda etapa, se seleccionaron los estudiantes dentro de las escuelas. Los procedimientos de muestreo en Puerto Rico no difirieron de los procedimientos en otras jurisdicciones, porque el propósito es incluir a Puerto Rico en la muestra de Estados Unidos en futuras administraciones de NAEP.

Para las administraciones de NAEP a modo de prueba en 2003 y 2005, se muestrearon aproximadamente 100 escuelas y 3,000 estudiantes por grado en escuelas públicas en Puerto Rico. Las escuelas privadas no participaron en las administraciones de NAEP a modo de prueba en Puerto Rico. La Tabla 5.1 presenta los tamaños de las muestras y las poblaciones meta para la evaluación de matemáticas de cuarto y octavo grado de 2003 y 2005 de estudiantes de escuelas públicas en Puerto Rico y en Estados Unidos. La información de Estados Unidos incluye sólo a estudiantes de escuelas públicas, aunque también participaron estudiantes de escuelas privadas.

Tabla 5.1
Tamaños de muestra y poblaciones meta de estudiantes de escuelas públicas en Puerto Rico y Estados Unidos para la evaluación NAEP de matemáticas, por grado y año

Año y grado	Puerto Rico		Estados Unidos	
	Tamaño de muestra de estudiantes	Población meta	Tamaño de muestra de estudiantes	Población meta
2003				
4º grado	3,000	48,000	191,400	3,603,000
8º grado	2,800	45,000	153,500	3,575,000
2005				
4º grado	2,800	42,000	168,900	3,745,000
8º grado	2,800	40,000	159,200	3,662,000

NOTA: Los tamaños de muestras de estudiantes están redondeados a la centésima más próxima y las poblaciones meta están redondeadas a la milésima más próxima.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Tasas de Participación de Escuelas y Estudiantes

Para reducir la posibilidad de estimados sesgados, NCES y la Junta Regidora establecieron normas para las tasas de participación que todas las jurisdicciones, incluyendo Puerto Rico, deben cumplir. NCES exige un análisis de la tendencia a no responder si la tasa de participación es inferior a 85 por ciento. La Junta Regidora exige una tasa de respuesta de 70 por ciento para fines de informes. En las evaluaciones de 2003 y 2005 de Puerto Rico se cumplieron ambos estándares de tasas de participación, tanto en cuarto como en octavo grado. La Tabla 5.2 proporciona las tasas de participación de escuelas y la Tabla 5.3 proporciona las tasas de participación de estudiantes.

Tabla 5.2

Número y porcentaje de escuelas públicas que participaron en la evaluación NAEP de matemáticas en Puerto Rico y Estados Unidos, por grado y año

Año y grado	Puerto Rico		Estados Unidos	
	Número de escuelas	Porcentaje ponderado	Número de escuelas	Porcentaje ponderado
2003				
4º grado	110	100	6,910	99.8
8º grado	100	100	5,530	99.6
2005				
4º grado	110	100	8,700	99.6
8º grado	110	100	6,460	99.5

NOTA: El número de escuelas está redondeado a la décima más próxima.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP). Evaluación de Matemáticas de 2003 y 2005.

Tabla 5.3

Número y porcentaje de estudiantes de escuelas públicas que participaron en la evaluación NAEP de matemáticas en Puerto Rico y Estados Unidos, por grado y año

Año y grado	Puerto Rico		Estados Unidos	
	Número de estudiantes	Porcentaje ponderado	Número de estudiantes	Porcentaje ponderado
2003				
4º grado	3,000	94	184,300	94
8º grado	2,800	92	147,600	91
2005				
4º grado	2,800	95	163,000	94
8º grado	2,800	93	152,800	91

NOTA: El número de estudiantes está redondeado a la centésima más cercana.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005

Referencias

- Allen, N.J., Donoghue, J.R., y Schoeps, T.L. (2001). *The NAEP 1998 Technical Report*. (NCES 2001-509). U.S. Department of Education, Office of Educational Research and Improvement. Washington, DC: National Center for Education Statistics.
- Allen, N., Jenkins, F., Kulick, E., y Zelenak, C.A. (1997). *Technical report of the NAEP 1996 State Assessment Program in Mathematics*. (NCES-1997-951). U. S. Department of Education, Office of Education Research and Improvement. Washington, DC: National Center for Education Statistics.
- Embretson, S. y Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hambleton, R.K., Merenda, P.F., y Spielberger, C.D. (Eds.) (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R.K., Swaminathan, H., y Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.
- Holland, P.W. y Wainer, H. (Eds.) (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F.M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- No Child Left Behind Act of 2001, Pub. L. 107-110, 115 Stat. 1425 (2002).
- The Elementary and Secondary Education Act (ESEA) (Pub. L. 89-10. 79Stat. 77, 20 U.S.C. ch.70)

Apéndice

Tabla A-1

Distribución porcentual del conjunto total y de los conjuntos de preguntas eliminadas y conservadas en cuarto grado para la evaluación NAEP de matemáticas, por año y característica de la pregunta

Característica de la pregunta	Eliminada	Conservada	Total
2003¹			
Área de contenido			
Numeración y operación	32	47	42
Medición	19	17	18
Geometría	18	14	15
Análisis de datos y probabilidad	11	11	11
Álgebra y funciones	21	11	15
Tipo de pregunta*			
Selección múltiple	79	57	64
Respuesta construida corta	18	39	32
Respuesta construida extendida	4	5	4
Habilidad en matemáticas			
Comprensión de conceptos	46	37	40
Resolución de problemas	40	34	36
Conocimiento de procedimientos	14	29	24
2005¹			
Área de contenido			
Numeración y operación	43	41	42
Medición	24	19	20
Geometría	5	17	14
Análisis de datos y probabilidad	11	11	11
Álgebra	16	12	13
Tipo de pregunta*			
Selección múltiple	92	56	64
Respuesta construida corta	3	40	32
Respuesta construida extendida	5	4	4
Complejidad matemática			
Baja	65	70	69
Moderada	32	30	30
Elevada	3	0	1

* La distribución de las preguntas eliminadas difiere significativamente ($p < .05$) de la distribución de las preguntas conservadas.

¹ El marco teórico de NAEP de matemáticas utilizado en 2005 difirió del marco teórico utilizado en 2003. En inglés, los nombres de algunas áreas de contenido cambiaron. La habilidad matemática se evaluó en 2003. La complejidad matemática se evaluó en 2005.

NOTA: La suma de los datos desglosados puede no corresponder al total debido al redondeo.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Tabla A-2

Distribución porcentual del conjunto total y de los conjuntos de preguntas eliminadas y conservadas en octavo grado para la evaluación NAEP de matemáticas, por año y característica de la pregunta

Característica de la pregunta	Eliminada	Conservada	Total
2003¹			
Área de contenido			
Numeración y operación	24	27	26
Medición	16	15	15
Geometría	18	19	18
Análisis de datos y probabilidad	11	17	15
Álgebra y funciones	31	22	25
Tipo de pregunta*			
Selección múltiple	73	63	66
Respuesta construida corta	26	31	30
Respuesta construida extendida	2	6	5
Habilidad en matemáticas			
Comprensión de conceptos	45	34	37
Resolución de problemas	31	35	34
Conocimiento de procedimientos	24	31	29
2005¹			
Área de contenido			
Numeración y operación	23	28	26
Medición	12	17	16
Geometría	21	21	21
Análisis de datos y probabilidad	16	13	14
Álgebra	28	21	23
Tipo de pregunta*			
Selección múltiple	75	65	69
Respuesta construida corta	19	31	28
Respuesta construida extendida	5	3	4
Complejidad matemática			
Baja	61	64	63
Moderada	39	33	35
Elevada	0	2	2

¹ El marco teórico de NAEP de matemáticas utilizado en 2005 difirió del marco teórico utilizado en 2003. En inglés, los nombres de algunas áreas de contenido cambiaron. La habilidad matemática se evaluó en 2003. La complejidad matemática se evaluó en 2005.

NOTA: La suma de los datos desglosados puede no corresponder al total debido al redondeo.

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

Tabla A-3

Valores máximos y mínimos de la puntuación promedio de las preguntas para el conjunto total y el conjunto reducido de NAEP de matemáticas, por año, grado y jurisdicción

Jurisdicción	Conjunto total de preguntas		Conjunto reducido de preguntas	
	Puntuación promedio mínima	Puntuación promedio máxima	Puntuación promedio mínima	Puntuación promedio máxima
2003				
4º grado				
Puerto Rico	.00	.79	.01	.75
Estados Unidos	.13	.95	.18	.95
8º grado				
Puerto Rico	.00	.89	.00	.89
Estados Unidos	.12	.94	.14	.94
2005				
4º grado				
Puerto Rico	.02	.79	.02	.79
Estados Unidos	.13	.96	.13	.96
8º grado				
Puerto Rico	.01	.90	.02	.90
Estados Unidos	.07	.94	.11	.94

FUENTE: Departamento de Educación de EE. UU., Instituto de Ciencias de la Educación, Centro Nacional para Estadísticas de la Educación, Evaluación Nacional del Progreso Educativo (NAEP), Evaluación de Matemáticas de 2003 y 2005.

DEPARTAMENTO DE EDUCACIÓN DE EE.UU.

La Evaluación Nacional del Progreso Educativo (NAEP, por sus siglas en inglés) es un proyecto requerido por mandato del Congreso y patrocinado por el Departamento de Educación de EE. UU. El Centro Nacional para Estadísticas de la Educación (NCES, por sus siglas en inglés), un departamento del Instituto de Ciencias de la Educación, administra NAEP. Por ley, el Comisionado de Estadísticas de la Educación es el responsable de llevar a cabo el proyecto NAEP.

Margaret Spellings
Secretaria
Departamento de Educación
de EE. UU.

Grover J. Whitehurst
Director
Instituto de Ciencias de la
Educación

Mark Schneider
Comisionado
Centro Nacional para
Estadísticas de la Educación

Peggy Carr
Comisionado Asociado
Centro Nacional para
Estadísticas de la Educación

LA LIBRETA DE CALIFICACIONES DE LA NACIÓN

Informe Técnico de la Evaluación NAEP de Matemáticas en Puerto Rico: Enfoque en Asuntos Estadísticos

Diciembre de 2007

CITA SUGERIDA

Baxter, G.P., Ahmed, S. Sikali, E. Waits, T., Sloan, M., y Salvucci, S. *Technical Report of the NAEP Mathematics Assessment in Puerto Rico: Focus on Statistical Issues (NCES 2008-477).*
U.S. Department of Education, National Center for Education Statistics, Washington, D.C.:
U.S. Government Printing Office.

CONTACTO RESPECTO AL CONTENIDO

Emmanuel Sikali • 202-502-7419 • emmanuel.sikali@ed.gov

MÁS INFORMACIÓN

- El sitio Web donde se publican estos informes es <http://nationsreportcard.gov>
- El sitio Web del catálogo electrónico (en inglés) de NCES es <http://nces.ed.gov/pubsearch>.
- Para información sobre pedidos escriba a:
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398
- o llame gratuitamente al 1-877-4ED-Pubs
- o pídalo por Internet al sitio Web <http://www.edpubs.org>



**“NUESTRA MISIÓN ES ASEGURAR EL ACCESO POR IGUAL A LA EDUCACIÓN
Y PROMOVER LA EXCELENCIA EDUCATIVA EN TODA LA NACIÓN.”**

www.ed.gov