

Guadalupe Carmona
The University of Texas at Austin
Lucía Monroy Cazorla
Miguel Herrera
Ingrid García
Centro Nacional de Evaluación para la Educación Superior (CENEVAL)

Prácticas docentes de los maestros de matemáticas y pruebas estandarizadas a gran escala. Evaluación formativa y las pruebas estandarizadas en México

Sinopsis

El propósito del estudio es conocer las prácticas docentes (en particular las relacionadas con la evaluación formativa) de los maestros de matemáticas de secundaria, y explorar cómo se relacionan con los resultados de los alumnos en el Exani I. Se revisaron las respuestas de 478,032 estudiantes en el examen y en el cuestionario de contexto. Las prácticas docentes fueron caracterizadas de acuerdo al marco conceptual de Ruiz-Primo (2004, 2006), del cual se elaboraron 3 escalas: epistémico, conceptual y social. Se encontraron bajas correlaciones entre las prácticas docentes y los puntajes en el examen. Estos resultados se discuten en términos de la congruencia que existe entre las prácticas docentes y el tipo de reactivos que se aplican en las evaluaciones a gran escala. En la medida en que se tenga claro cómo se relacionan, se podrán establecer conclusiones más válidas de los resultados de las pruebas a gran escala y de las inferencias y limitaciones que se puedan hacer con estos resultados.

Abstract

The purpose of this study was to determine the teaching practices (in particular those related to formative assessment) of middle school math, and if these are related with their students' test results in the EXANI I. We reviewed the responses of 478,032 students in EXANI I and in the background questionnaire. Teaching practices were characterized according to the conceptual framework of Ruiz-Primo (2004, 2006), from which were developed at 3 scales: epistemological, conceptual and social. Low correlations were found between teacher practices and the scores on the test. These results are discussed in terms of the congruence between teaching practices and the type of items that are used in large-scale assessments. To the extent that this relationship becomes more clear, it may provide more valid conclusions, inferences, and limitations when results of large-scale assessments are used.

Términos clave: Investigación, Prácticas de enseñanza de las matemáticas, Evaluación formativa, EXANI I, México.

Keywords: Research, Mathematics instruction, Evaluation, Mexico.

Fecha de recepción: Febrero 2011

Fecha de aprobación: Mayo 2011

Introducción

En los últimos 25 años a nivel internacional, las evaluaciones a gran escala se han convertido en uno de los principales factores sobre los cuales se fundamentan las decisiones educativas de alto impacto. Este tipo de evaluaciones son realizadas con dos propósitos: la selección de sustentantes y la evaluación de la calidad de la educación. Para este último objetivo existe el supuesto de que las evaluaciones a gran escala se pueden utilizar para hacer inferencias sobre el aprendizaje y la enseñanza de los estudiantes. Sin embargo, en muchas ocasiones, los estudios han dado cuenta de una desconexión entre las calificaciones de los estudiantes en las evaluaciones a gran escala y el desempeño en el aula (NRC, 2001; Carmona, 2006, en revisión).

Debido a esta situación existe un amplio consenso de la necesidad de tener una mayor integración entre la evaluación y la instrucción (NRC, 2001). Por lo que se ha propuesto que las evaluaciones deben ser más sensibles a la instrucción (Popham, 2007, 2008; Stroup, 2009; Wiliam 2007, 2008), lo que posibilitaría realizar mejores inferencias sobre el aprendizaje de los estudiantes y la calidad de la enseñanza o instrucción.

El Centro Nacional para la Evaluación de la Educación Superior (Ceneval) diseña y administra pruebas y evaluaciones con fines académicos, principalmente para las escuelas del nivel medio superior y las universidades. Una de estas pruebas es el Examen Nacional de Ingreso a la Educación Media Superior (EXANI I), una prueba estandarizada de alto impacto que se implementa a escala nacional en México para seleccionar a los sustentantes que ingresan al nivel medio superior.

Con la intención de desarrollar evaluaciones a gran escala que sean más sensibles a la instrucción, el Ceneval incorporó en 2010, 21 preguntas en el cuestionario de contexto para estudiantes del EXANI I. Estas preguntas describen las prácticas de enseñanza que son utilizadas por los profesores de matemáticas del último grado de secundaria. Algunas de estas prácticas comprenden una evaluación formativa en el aula, este tipo de evaluación considera a las prácticas de evaluación y a la instrucción como dos elementos muy relacionados entre sí (NRC, 2001). De acuerdo a Black y Wiliam (1998a, 1998b), la evaluación formativa se refiere a las actividades que proporcionan evidencias de la comprensión de los estudiantes y que son utilizadas por los profesores para tomar decisiones sobre su propia práctica docente, estos autores reportan que de todas las prácticas de enseñanza, la evaluación formativa es la que tiene un mayor efecto en el rendimiento de los estudiantes. Por su parte, Bell y Cowie (2001) la definen como recolectar, interpretar y actuar sobre la información acerca de lo que aprenden los estudiantes. En ambos casos a partir de la información se les proporciona retroalimentación a los estudiantes y se modifican los modos de enseñanza. Es necesario estudiar a mayor detalle las prácticas de evaluación formativa que los maestros implementan en sus aulas para recabar información sobre la comprensión de los estudiantes del contenido abordado en clase, debido a que la evaluación formativa es esencial para una buena enseñanza y aprendizaje (Black, 1993; Black y Wiliam, 1998a). Si los profesores no tienen información sobre el avance de la comprensión de sus estudiantes, sus esfuerzos para ayudarlos a mejorar su aprendizaje son limitados.

De acuerdo a Ruiz-Primo y Furtak (2006, 2004), la evaluación formativa se puede caracterizar centrándola en tres áreas: marcos epistémicos, estructuras conceptuales y procesos sociales. Las estructuras epistémicas son los marcos de conocimiento que implican las normas y criterios utilizados para desarrollar y/o juzgar lo que se puede entender como

científico (por ejemplo, experimentos, hipótesis o explicaciones). Estos marcos también abarcan el desarrollo de los criterios para emitir juicios acerca de los productos de una investigación. Las estructuras conceptuales implican la comprensión profunda de los conceptos y principios de los grandes esquemas científicos para aplicarlos de forma eficaz en las situaciones apropiadas. Los procesos sociales se refieren a los marcos de participación de los estudiantes, se centran en cómo el conocimiento se comunica, representa, y argumenta. El propósito de este estudio es caracterizar la enseñanza de las matemáticas en el aula, de acuerdo al marco establecido por Ruiz-Primo (2004, 2006), a través de las percepciones de los estudiantes que identifiquen si sus profesores conocen, reconocen y utilizan procedimientos y técnicas que fomenten la evaluación formativa. Adicionalmente se quiere conocer la posible relación de estas prácticas con las calificaciones de los estudiantes en las secciones de matemáticas y lógica matemática del EXANI I.

Este estudio se guió por las siguientes preguntas:

- a) ¿Cuáles son las prácticas docentes más utilizadas por los profesores de matemáticas en el último grado de secundaria?
- b) ¿Qué prácticas de enseñanza en las clases de matemáticas, se encuentran asociadas a un mejor puntaje en las secciones de razonamiento lógico matemático y de matemáticas?
- c) ¿Es posible desarrollar escalas psicométricas para medir los dominios que comprenden la evaluación formativa?

Método

Participantes

Los participantes fueron 478,032 sustentantes mexicanos de todo el país que deseaban ingresar al nivel medio superior y contestaron de enero a julio del 2010 el EXANI I y el cuestionario de contexto.

Herramientas

- EXANI I

El EXANI I es una prueba de opción múltiple que mide las aptitudes académicas y proporciona información sobre las habilidades intelectuales y los conocimientos disciplinarios de los estudiantes para su ingreso en el nivel medio superior. En la parte de selección, el EXANI I cuenta con cuatro secciones: razonamiento verbal, español, razonamiento lógico matemático y matemáticas. El EXANI I incluye un total de 80 reactivos de opción múltiple, 20 por cada sección. Cada reactivo cuenta con 4 opciones de respuesta.

Los análisis desarrollados se centraron en los 40 reactivos, correspondientes a las secciones de razonamiento lógico matemático y de matemáticas.

- Cuestionario de contexto

Además del EXANI I, cada estudiante contestó un cuestionario de contexto de 120 preguntas. Este cuestionario mide constructos como: capital económico y capital cultural, entre otros. En 2010 fue incorporada una nueva sección que incluye 21 preguntas sobre las prácticas de los profesores de matemáticas, distribuidas en tres escalas: epistémica, conceptual y social. En la tabla 1 se muestra la distribución de las preguntas por cada escala.

Tabla 1
Distribución de las preguntas por escala.

Escala	Número de preguntas
Epistémica	13
Conceptual	3
Social	5

En la tabla 2 se muestra la lista de las 21 preguntas, cada una se presentó con cuatro opciones de respuesta: nunca, pocas veces, frecuentemente, y siempre.

Tabla 2
Preguntas de las prácticas de enseñanza organizadas por escala.

Escala	Número	Pregunta
Epistémica	1	Indica con qué frecuencia tu maestro consideraba las experiencias de los alumnos para explicar los temas del curso
	2	Solicitaba que los alumnos evaluaran si el procedimiento que él había utilizado era el más correcto o apropiado para un ejercicio o problema
	3	Indica con qué frecuencia tu maestro mostraba las aplicaciones de los conceptos aprendidos
	4	Indica con qué frecuencia tu maestro resolvía ejercicios en clase
	5	Indica con qué frecuencia tu maestro relacionaba tus experiencias con los temas del curso
	6	Indica con qué frecuencia tu maestro dejaba ejercicios con una sola forma de resolverlos y una sola respuesta correcta
	7	Indica con qué frecuencia tu maestro dejaba ejercicios con varias formas de resolverlos pero una sola respuesta correcta
	8	Indica con qué frecuencia tu maestro dejaba ejercicios con varias formas de resolverlos y varias respuestas correctas
	9	Indica con qué frecuencia tu maestro ante una respuesta incorrecta de un alumno corregía inmediatamente el error y daba la respuesta correcta
	10	Indica con qué frecuencia tu maestro ante una respuesta incorrecta de un alumno realizaba otras preguntas relacionadas con el tema a todo el grupo hasta encontrar la respuesta correcta
	11	Indica con qué frecuencia tu maestro ante una respuesta incorrecta de un alumno preguntaba a otro alumno que pudiera dar la respuesta correcta
	12	Indica con qué frecuencia tu maestro pedía a los alumnos que encontrarán diferentes maneras de resolver un problema o ejercicio
	13	Indica con qué frecuencia tu maestro explicaba las respuestas correctas y también las incorrectas

Escala	Número	Pregunta
Conceptual	14	Indica con qué frecuencia tu maestro utilizaba gráficas como esquemas, cuadros sinópticos, mapas conceptuales, para explicar los temas del curso
	15	Indica con qué frecuencia tu maestro aclaraba dudas al terminar de revisar un tema
	16	Indica con qué frecuencia tu maestro pedía que se proporcionaran ejemplos de los temas expuestos por él
Social	17	Indica con qué frecuencia tu maestro pedía a los alumnos que participaran durante la clase (con la propuesta de algún ejemplo, la resolución de un problema, la explicación de un tema, etc.)
	18	Indica con qué frecuencia tu maestro permitía que todos los estudiantes tuvieran la oportunidad de participar
	19	Indica con qué frecuencia tu maestro facilitaba el trabajo individual
	20	Indica con qué frecuencia tu maestro facilitaba el trabajo en equipo
	21	Indica con qué frecuencia tu maestro facilitaba el trabajo de todo el grupo

Procedimiento

Durante el año 2010 el Ceneval aplicó el EXANI I a los estudiantes que querían ingresar al nivel medio superior. Adicionalmente antes de presentar el examen los estudiantes respondieron el cuestionario de contexto.

Resultados

Para responder las preguntas de investigación, se llevaron a cabo los siguientes análisis estadísticos: En primer lugar, se obtuvieron las frecuencias de las 21 preguntas incluidas en el cuestionario de contexto, con lo que se determinó cuáles son las prácticas de enseñanza de matemáticas más comunes en el aula de acuerdo a la percepción de los estudiantes. En segundo lugar, se obtuvieron las características psicométricas de las escalas epistemológica, conceptual y social. En tercer lugar, se calcularon las correlaciones entre los puntajes de las escalas con los puntajes en matemáticas y en razonamiento lógico matemático. En cuarto lugar, se calculó la correlación entre cada una de las preguntas de las prácticas de enseñanza de matemáticas con cada uno de los reactivos de las secciones de matemáticas y de razonamiento lógico matemático.

Análisis de frecuencias

Las respuestas de los sustentantes se dicotomizaron, ya que las frecuencias de la pregunta 1 del cuestionario se distribuyeron formando una U inversa (ver Tabla 3). Esto mismo ocurrió con la mayoría de las demás preguntas del cuestionario, por lo cual, se decidió que era conveniente dicotomizarlas, asignado el valor de 1 para las opciones de siempre y frecuentemente y valor de 0 para las opciones de nunca y pocas veces.

Tabla 3

Frecuencias de respuesta de la pregunta 1 de las prácticas docentes de los maestros en matemáticas.

	Frecuencia	Porcentaje
Nunca	39640	8.3
Pocas veces	177702	37.2
Frecuentemente	157728	33.0
Siempre	63174	13.2
Total	438244	91.7
Datos faltantes	39788	8.3

Las figuras 1, 2 y 3 muestran para cada una de las escalas (epistémico, conceptual y social) el porcentaje de estudiantes que reportaron que su profesor de matemáticas utilizaba dichas prácticas docentes frecuentemente o siempre.

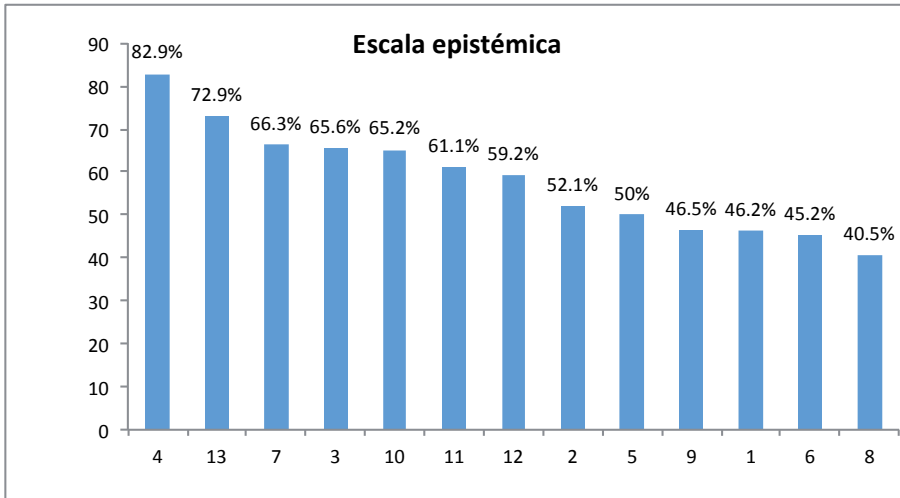


Figura 1

Porcentaje de estudiantes que reportaron que los maestros utilizaban cada práctica docente frecuentemente o siempre en la escala epistémica.

En la figura 1, se observa que las prácticas de enseñanza más frecuentes de la escala epistémica corresponden a las preguntas 4, 13 y 7. Mientras que las menos frecuentes están descritas en las preguntas 1, 6 y 8.

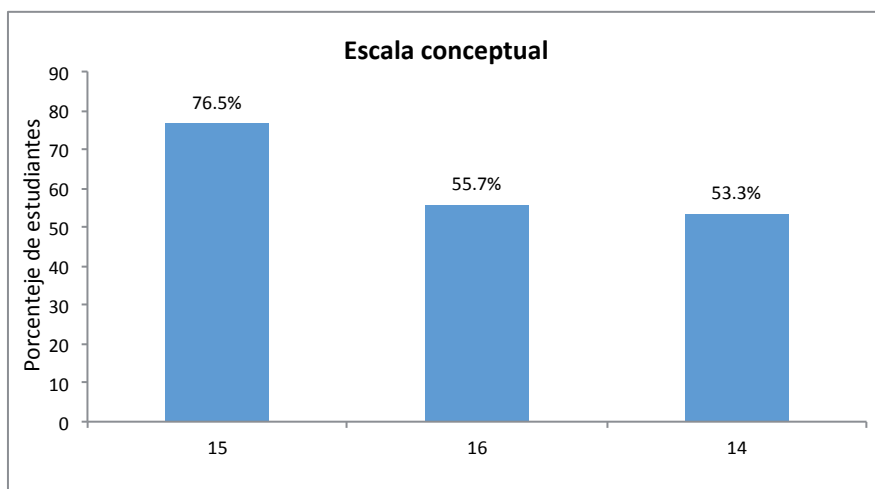


Figura 2
Porcentaje de estudiantes que reportaron que los maestros utilizaban cada práctica docente frecuentemente o siempre en la escala conceptual.

En la figura 2 se observa que en la escala conceptual, la práctica docente más frecuente fue la correspondiente a la pregunta 15, seguida por las de las preguntas 16 y 14.

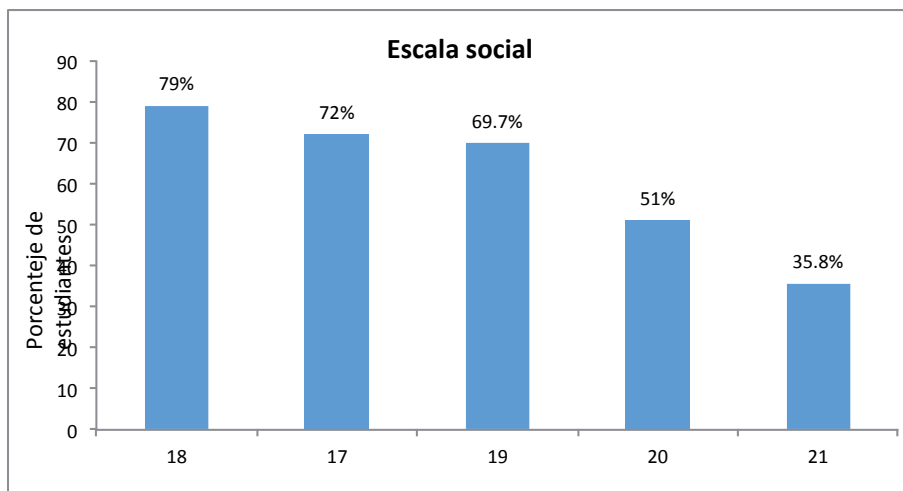


Figura 3
Porcentaje de estudiantes que reportaron que los maestros utilizaban cada práctica docente frecuentemente o siempre en la escala social.

En la figura 3 se muestra que en la escala social, la práctica con mayor frecuencia fue la pregunta 18.

De estos resultados podemos describir un perfil común de las actividades docentes en el salón de clase de matemáticas del último año de secundaria en México, de acuerdo a la percepción de los estudiantes. Esta caracterización corresponde a una práctica centrada en el maestro y los estudiantes en general trabajan de manera individual. El maestro es quien resuelve los ejercicios en clases, seleccionado aquellos que tienen una única respuesta correcta. El maestro explica las respuestas correctas e incorrectas, y permite que los estudiantes resuelvan los problemas de diferentes maneras, abriendo la oportunidad a que los estudiantes participen a través de explicaciones de la resolución de los problemas.

Propiedades psicométricas de las escalas y sus correlaciones con las secciones de matemáticas y de razonamiento lógico matemático

Las tres escalas que describen las prácticas de enseñanza de matemáticas se analizaron mediante el modelo de Rasch de crédito parcial (PCM; Masters, 1982). Este modelo fue utilizado para estimar las propiedades psicométricas de las escalas, por su utilidad para el análisis de datos politómicos.

En la tabla 4 se presentan las características psicométricas de las escalas. En la primera y segunda columna se muestran las preguntas de acuerdo a la escala que pertenecen. En la tercera y cuarta columna contienen el ajuste de cada pregunta al modelo de crédito parcial. Las siguientes dos columnas señalan el alfa de Cronbach y la varianza explicada por un único factor. Por último, las dos últimas columnas contienen la correlación entre el puntaje de la escala con el porcentaje de aciertos de las secciones de razonamiento lógico matemático y de matemáticas.

Tabla 4
Características psicométricas de las escalas y su correlación con el puntaje en las secciones de razonamiento lógico matemático y matemáticas.

Escala	Pregunta	Ajuste al modelo de crédito parcial		Alfa de Cronbach	Varianza explicada por un único factor	Correlación con las secciones del EXANI I	
		INFIT	OUTFIT			Razonamiento lógico matemático	Matemáticas
Epistémica	7	1.12	1.24	0.76	56.6%	0.072	0.087
	10	1.06	1.13				
	12	1.08	1.12				
	4	1.03	1.05				
	13	0.99	1.01				
	5	0.97	0.97				
	1	0.96	0.96				
	2	0.92	0.89				
	3	0.85	0.78				
	17	0.91	0.91				
Social	21	1.06	1.37	0.43	62.4%	0.027	0.041
	20	0.99	1.09				
	18	0.96	0.92				
Conceptual	12	1.05	1.09	0.52	47.1%		
	16	1.03	1.08				
	14	0.92	0.90				

Vale la pena señalar que el ajuste al modelo de crédito parcial (tercera y cuarta columna) de la mayoría de las preguntas se encuentran dentro del rango aceptable (entre 0.7 y 1.3). También se observa que sólo la escala epistemológica obtuvo un valor satisfactorio en el Alfa de Cronbach calculada, mientras que la consistencia interna de las escalas sociales y conceptuales es baja, aunque esto podría ser consecuencia de un número reducido de preguntas, por lo que es recomendable que el número se incremente para versiones poste-

riores del cuestionario. Por último se observa que los valores de correlación encontrados para cada escala fueron bajos, tanto con el razonamiento lógico matemático como con matemáticas.

Correlaciones entre las prácticas docentes y los resultados en las secciones de matemáticas y de razonamiento lógico matemático

Adicionalmente también se llevaron a cabo análisis de correlación entre cada una de las prácticas docentes (dividas por escala) con los resultados en las secciones de razonamiento lógico matemático y de matemáticas, respectivamente. Los resultados se presentan en las figuras 4, 5, 6, 7, 8 y 9. Todas las correlaciones fueron significativas con un nivel de alfa de 0.05.

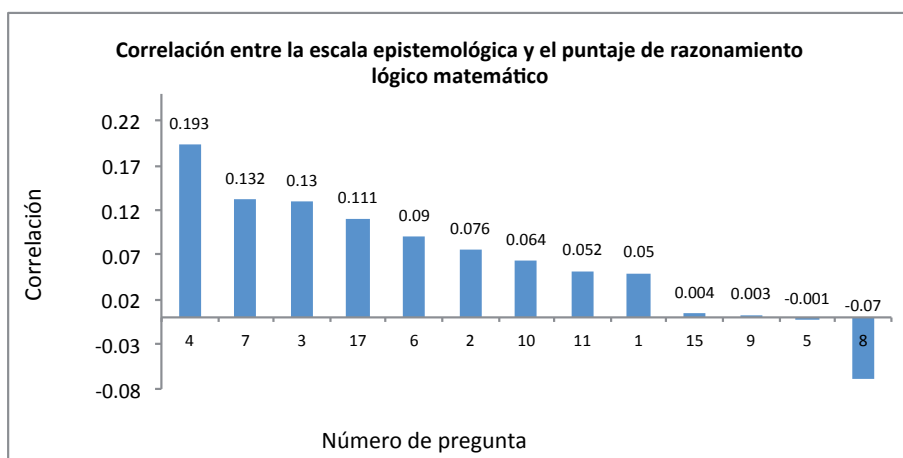


Figura 4

Correlaciones entre prácticas docentes de la escala epistémica y los resultados de razonamiento lógico matemático.

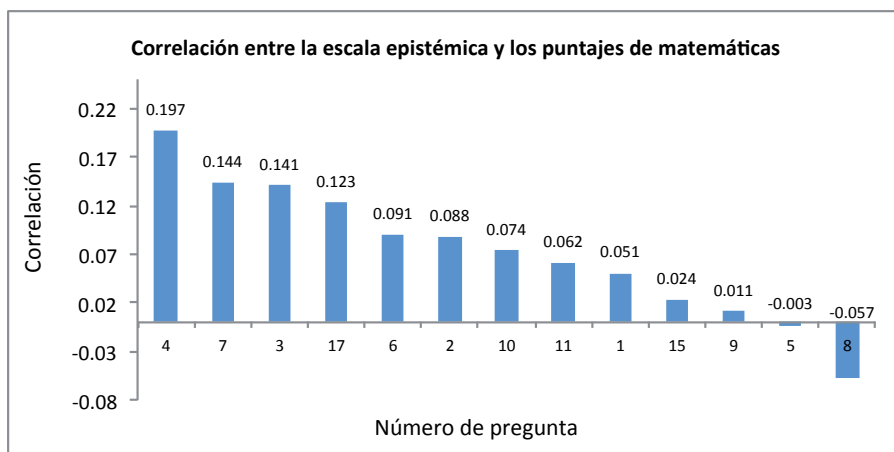


Figura 5

Correlaciones entre prácticas docentes de la escala epistémica y los resultados de matemáticas.

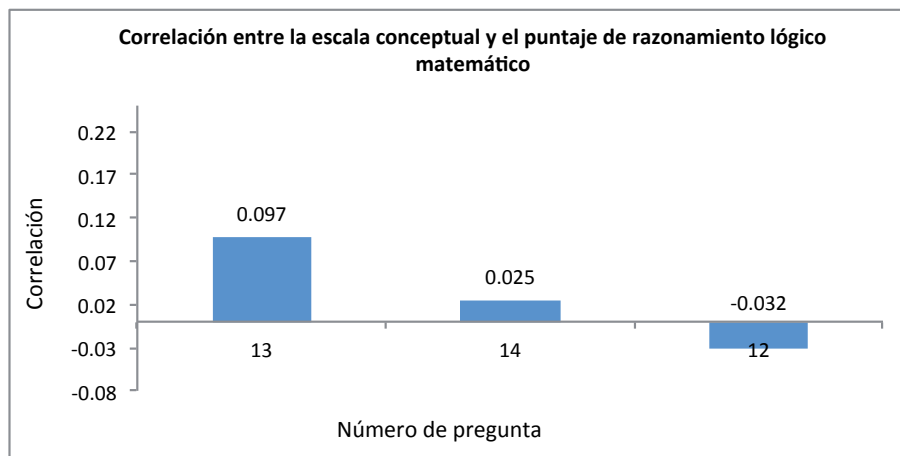


Figura 6
Correlaciones entre prácticas docentes de la escala conceptual y los resultados de razonamiento lógico matemático.

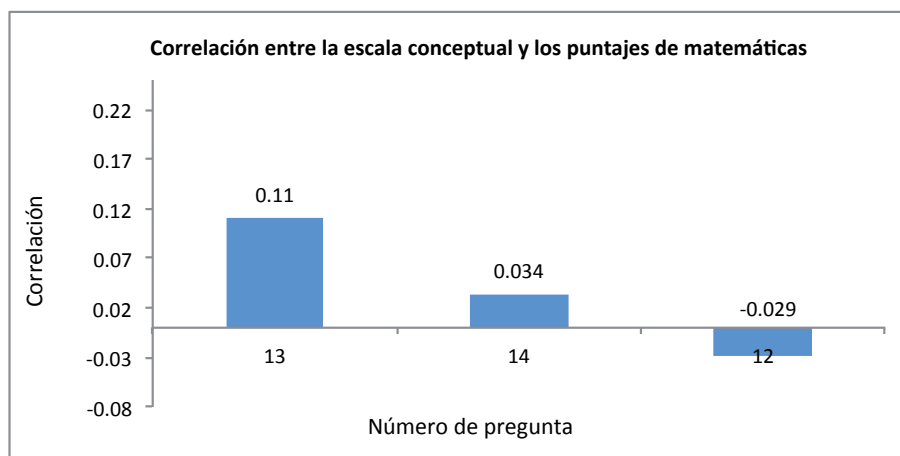


Figura 7
Correlaciones entre prácticas docentes de la escala conceptual y los resultados de matemáticas.

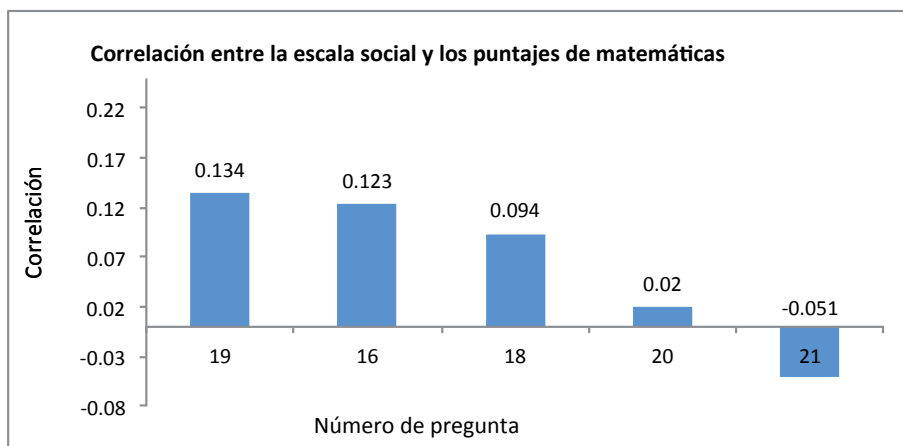


Figura 8
Correlaciones entre prácticas docentes de la escala social y los resultados de razonamiento lógico matemático.

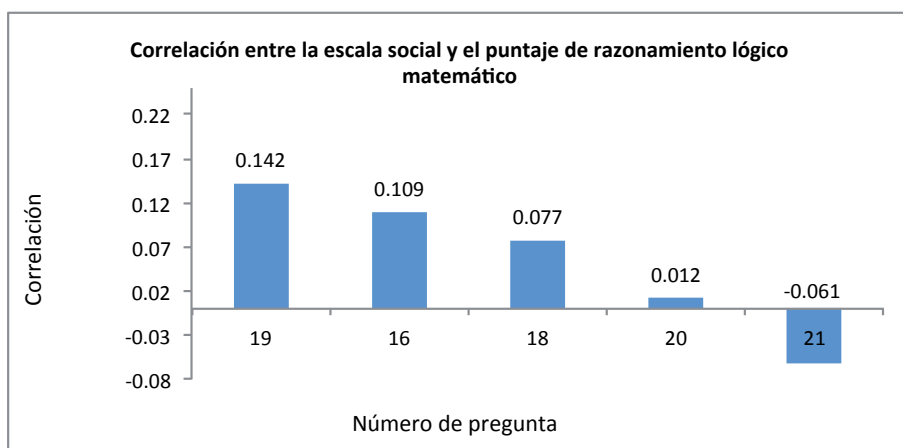


Figura 9
Correlaciones entre prácticas docentes de la escala social y los resultados de matemáticas.

Todas las correlaciones fueron muy bajas (entre -0.057 y 0.197). De los resultados presentados en estas seis figuras, las prácticas docentes con mayores valores de correlación fueron las correspondientes a las preguntas: 4, 7 y 3. Una observación interesante es la consistencia que existe entre los valores de las correlaciones de razonamiento lógico matemático y matemáticas, ya que es prácticamente una correspondencia uno a uno para cada pregunta.

Correlaciones entre las prácticas docentes y reactivos individuales de matemáticas y de razonamiento lógico matemático

Se realizó un análisis de correlación de cada uno de los reactivos de las seccio-

nes de matemáticas y de razonamiento lógico matemático, con cada una de las prácticas docentes. Por limitaciones de espacio, no se presenta la matriz completa de correlaciones, sino sólo casos representativos. Se encontró que los valores de las correlaciones más altos fueron entre 0.10 y 0.15 significativas con un nivel de alfa de 0.05, sin embargo, en la mayoría de los casos las correlaciones fueron más bajas (entre 0.05 y 0.01). Con estos valores se realizó un análisis para identificar las características de aquellos reactivos con que se obtuvieron valores altos o bajos. Los resultados indican que el tipo de actividad cognitiva que requieren los sustentantes para resolver cada uno de estos reactivos es cualitativamente distinto. Para ilustrar mejor estos resultados, se presentan los dos ejemplos siguientes.

En particular, un reactivo con una correlación más alta es uno en el que el estudiante tenía que encontrar el valor de la siguiente expresión:

Reactivo A

$$5x + 3y, \text{ si } x = 3, y = 2$$

En contraste, uno de los reactivos que obtuvo en general una baja correlación fue el siguiente:

Reactivo B

La masa en kilogramos de la Tierra es $5,97 \times 10^{24}$ y la de Saturno es $5,68 \times 10^{26}$. Aproximadamente, ¿cuántas veces es mayor la masa de Saturno en comparación con la de la Tierra?

Contrastando la actividad cognitiva requerida por parte del sustentate, se puede ver que el reactivo B requiere una actividad mucho mayor que el reactivo A. El reactivo A se puede resolver a través de la aplicación de un procedimiento automatizado, sustituyendo los valores correspondientes y realizando las operaciones indicadas de multiplicación y suma. En contraste, el reactivo B requiere que los estudiantes puedan interpretar la notación científica y poder cuantificar ambas magnitudes para determinar cuántas veces es mayor la masa de la Tierra que la de Saturno. Por tanto, requiere el manejo de conceptos como poder comparar de dos magnitudes de gran dimensión y escritas en notación científica. En la tabla 5 se muestra los coeficientes de correlación entre los reactivos A y B con las prácticas docentes con las que tuvieron valores más altos y más bajos.

Tabla 5
Coeficientes de correlación entre los reactivos A y B con las prácticas docentes.

Pregunta	Valores de correlación más altos	
	A	B
4	0.1550	0.0533
3	0.1271	0.0389
19	0.1081	0.0352
13	0.1247	0.0254
7	0.1061	0.0240
Pregunta	Valores de correlación más bajos	
	A	B
9	-0.0024	-0.0007
14	-0.0251	-0.0072
8	-0.0195	-0.0239
21	-0.0261	-0.0145

Al observar esta tabla, se puede ver que la práctica docente correspondiente a la pregunta 4 “el profesor resolvía ejercicios en clase” tiene una correlación tres veces mayor con el reactivo A que con el reactivo B.

Discusión y Conclusiones

Los resultados de este estudio muestran que las frecuencias de las prácticas docentes asociadas a la evaluación formativa fueron relativamente altas, sin embargo, se encontraron correlaciones bajas con los resultados de los estudiantes en las secciones de razonamiento lógico matemático y matemáticas. Lo anterior puede explicarse si se toma en cuenta que la mayoría de los reactivos requieren que los estudiantes apliquen procedimientos automatizados, por lo que si las prácticas asociadas a la evaluación formativa tienen un efecto positivo debe analizarse si los reactivos de una prueba permiten medir este efecto. Es decir, se debe analizar qué tanto los reactivos son sensibles a las prácticas de los docentes. Aunque esta explicación es tentativa, debe considerarse que es indispensable establecerse una teoría de los efectos de las prácticas docentes en distintas disciplinas como matemáticas o español y revisarse si las evaluaciones a gran escala miden esos efectos. En otras palabras debe tenerse una mayor claridad de la relación entre el rendimiento en las pruebas y la enseñanza. Sólo así se podrán establecer conclusiones de los resultados de las pruebas a gran escala que sirvan para dar recomendaciones sobre políticas educativas y evaluar la calidad de la educación con un mayor grado de validez y certidumbre. Actualmente al carecer de una teoría sólida que explique esta situación, puede ocurrir que si un maestro aplica prácticas docentes “recomendadas”, podría no tener ningún efecto si los reactivos de una prueba no miden los beneficios asociados a esa práctica. Es también necesario desarrollar un modelo que pueda categorizar los reactivos en un continuo de complejidad cognitiva, así se podría conocer si se está aplicando un reactivo complejo o uno más simple y se podría analizar el impacto de las prácticas docentes de acuerdo a los reactivos presentados. Por otra parte también es importante realizar más investigación de campo para hacer un análisis más detallado sobre las prácticas docentes. Por ejemplo, en la pregunta: “¿Con qué frecuencia tu maestro muestra la aplicación de los conceptos?” es posible que un maestro sólo exponga en su clase que algún concepto es útil para un ingeniero, mientras que otro maestro muestre la relación conceptual entre los términos o campos de aplicación; pero en ambos casos, la respuesta del estudiante sería que el maestro muestra la aplicación de los conceptos, cuando en realidad los maestros hacen cosas cualitativamente distintas. Por lo que en la medida en que se identifiquen qué prácticas docentes tienen un mayor impacto en el aprendizaje de los alumnos y qué tipo de reactivos pueden evaluar esto, se podrá implementar en los exámenes a gran escala reactivos que sean sensibles a estas prácticas.

Referencias

- Bell, B., & Cowie, B. (2001). *Formative Assessment and Science Education*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-148.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314(7080), 572.

- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Carmona, G. (2006). The relationship between performance-based assessments and states' standardized testing. Individual presentation in the SIG for Large Scale Assessment. Annual Meeting of the American Educational Research Association: San Francisco, CA.
- Mann, H. B. & Whitney, D. R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". *Annals of Mathematical Statistics* 18 (1): 50-60.
- Masters, G. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174-174.
- NRC (2001). *Knowing what students know: The science and design of educational assessments*. Washington, D.C.: National Research Council.
- Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan*, 89(2), 146-150.
- Popham, W. J. (2008). Formative assessment: Seven stepping-stones to success. *Principal Leadership*, 9(1), 16-20.
- Ruiz-Primo, M.A. & Furtak, E.M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3 & 4), p. 237-263.
- Ruiz-Primo, M. A., & Furtak, E. M. (2004). *Informal assessment of students' understanding of scientific inquiry*. CSE: Technical Report 639. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing/ University of California, Los Angeles.
- Smith, A., Rush, R., Fallowfield, L., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8(1), 33.
- Stroup, W. M. (2009). What it means for mathematics tests to be insensitive to instruction. Plenary In L. C. Hart & C. D. Thomas (Eds.) *Proceedings of the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. [CD-ROM]. Eugene, OR: All Academic.
- William, D. (2007). Three practical, policy-focused procedures for determining an accountability test's instructional sensitivity: An index of sensitivity to instruction. Paper presented at the Annual meeting of the American Educational Research Association held at Chicago, IL.
- William, D. (2008). International comparisons and sensitivity to instruction. *Assessment in*



DATOS DE LOS AUTORES

Guadalupe Carmona
The University of Texas at Austin
lcarmona@mail.utexas.edu

Lucía Monroy Cazorla
Centro Nacional de Evaluación para la Educación Superior (CENEVAL)
lucia.monroy@ceneval.edu.mx

Miguel Herrera
Centro Nacional de Evaluación para la Educación Superior (CENEVAL)
miguel.herrera@ceneval.edu.mx

Ingrid García
Centro Nacional de Evaluación para la Educación Superior (CENEVAL)
ingrid.garcia@ceneval.edu.mx