



RELIEVE. Revista Electrónica de Investigación
y Evaluación Educativa

E-ISSN: 1134-4032

relieve@uv.es

Universitat de València
España

Lopez, Alexis A.

Alineación entre las evaluaciones externas y los estándares académicos: el caso de la prueba Saber
de Matemáticas en Colombia

RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa, vol. 19, núm. 2, julio-
diciembre, 2013, pp. 1-16

Universitat de València
Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=91631052003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

ALINEACIÓN ENTRE LAS EVALUACIONES EXTERNAS Y LOS ESTÁNDARES ACADÉMICOS: EL CASO DE LA PRUEBA SABER DE MATEMÁTICAS EN COLOMBIA

*[Alignment between standardized assessments and academic standards:
The case of the Saber Mathematics Test in Colombia]*

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

Lopez, Alexis A. (alopez@ets.org)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

In this study the concept of alignment is analyzed by presenting the results of a study that examines the degree of alignment between a standardized mathematics assessment and a set of academic standards. The alignment was examined using the Webb model and the results suggest that the degree of alignment between standardized assessments and academic content standards is not the most appropriate and therefore the results should be interpreted with great caution and stakeholders must be very careful with the type of decisions that are made based on the results of these types of assessments.

Keywords

Alignment, Standardized Tests, Academic Standards, Assessment of learning, Validity.

Resumen

En este estudio se analiza el concepto de alineación presentando los resultados de un estudio que examina el grado de alineación entre una evaluación estandarizada de matemáticas y unos estándares académicos. La alineación se examinó usando el modelo de Webb y los resultados sugieren que el grado de alineación entre las evaluaciones estandarizadas y los estándares académicos no es el más adecuado y por lo tanto los resultados se deben interpretar con mucha cautela y se debe poner mucho cuidado con el tipo de decisiones que se toman con base en los resultados.

Descriptores

Alineación, Pruebas estandarizadas, Estándares Académicos, Evaluación del aprendizaje, Validez.

La evaluación del aprendizaje es un aspecto fundamental en el proceso de enseñanza y aprendizaje, ya que los resultados se utilizan para obtener información importante sobre lo que los estudiantes conocen y saben hacer, y para tomar decisiones importantes que van a afectar sus vidas. Por lo tanto, es crítico que estas evaluaciones sean válidas y apropiadas según su propósito. Para que una evaluación sea válida, como mínimo, este tiene que estar alineado con lo que se espera que se enseñe en el aula de clase. Por alineación se entiende que hay acuerdo entre lo que se evalúa y lo que

está descrito en el currículo (Webb, 1997). Generalmente, lo que se enseña en clase está guiado por lo que dicen los estándares. Esta alineación entre los exámenes y los estándares ayuda a que los estudiantes aprendan lo que está estipulado en el currículo.

El concepto de alineación está intrínsecamente relacionado con el concepto de validez. La validez de un examen se evalúa teniendo en cuenta evidencias empíricas y justificaciones teóricas que sustentan la idoneidad de las interpretaciones y las decisiones que se toman con base en los resultados del mismo (Mes-

sick, 1989). Existen dos aspectos fundamentales que pueden invalidar las interpretaciones que se hacen con base en los resultados de un examen: 1) que el contenido y las competencias que se evalúan en el examen no estén bien representadas y 2) que el contenido y las competencias que se evalúan en el examen incluyan facetas o dimensiones irrelevantes (Messick, 1989). En el primer caso, el contenido del examen es muy restringido o limitado y no incluye facetas o dimensiones importantes que están descritas en los estándares. Por el contrario, en el segundo caso el contenido del examen es muy amplio y se evalúan contenidos o competencias que no están definidas en los estándares.

Según el grado de alineación que exista entre una evaluación y unos estándares, se pueden dar cuatro situaciones diferentes. En la primera situación, todo el contenido de la evaluación está alineado con los estándares, pero hay muchos contenidos de los estándares que no se están evaluando. En la segunda situación, todos los contenidos de los estándares están siendo evaluados, pero la evaluación también está midiendo muchos aspectos que no están enunciados en los estándares. En la tercera situación, solamente una parte de los estándares está siendo evaluada, pero la evaluación está midiendo algunos aspectos que no están enunciados en los estándares y mucho de los contenidos de los estándares no están siendo evaluados. Por último, tenemos la situación ideal donde todas las preguntas están evaluando una muestra representativa de todos los contenidos enunciados en los estándares. Esta es una situación difícil de alcanzar ya que se tienen muchas limitaciones. Por ejemplo, el número de contenidos en los estándares, el número de preguntas en la evaluación y la forma como se evalúa (el tipo de preguntas que se usan).

En el caso de muchas evaluaciones, generalmente el contenido está definido por los contenidos descritos en los estándares académicos. Los resultados de las evaluaciones indican qué tanto han aprendido los estudiantes

con relación a los contenidos de los estándares. Si las evaluaciones no están alineadas con los contenidos de los estándares, y por ende con lo que los profesores enseñan en el aula de clase, los resultados no son del todo válidos para determinar qué tanto se han alcanzado las expectativas que se tienen al terminar cada año académico. Por lo tanto, es muy importante determinar el grado de alineación que hay entre las evaluaciones estandarizadas y los estándares académicos, para determinar cómo se deben interpretar estos resultados y para entender qué tipo de decisiones se pueden tomar con base en los mismos.

Los estudios de alineación también pueden servir para identificar áreas problemáticas en las evaluaciones (López, Webb & Stansfield, 2006; Rothman, 2004). Los resultados nos permiten identificar vacíos en el contenido de una evaluación y nos ayudan a determinar si éste es lo suficientemente representativo de los estándares (Herman, 2004; Porter, 2002; Rothman, 2004; Wise, 2004).

El objetivo de este estudio es ilustrar el concepto de alineación determinando el nivel de alineación entre la Prueba Saber de Matemáticas y los Estándares Básicos de Competencias en Matemáticas colombianos. Para establecer el grado de alineación, se examinaron las siguientes preguntas de investigación:

1. ¿Cuál es el grado de alineación entre el contenido de la Prueba Saber de Matemáticas y el contenido de los estándares?
2. ¿Cuál es el grado de alineación entre la demanda cognitiva de las preguntas en la Prueba Saber de Matemáticas y la demanda cognitiva en los estándares?
3. ¿Cuál es el grado de alineación entre la cobertura de las preguntas en la Prueba Saber de Matemáticas y la cobertura de los estándares?
4. ¿Cuál es el grado de alineación entre la forma como están distribuidas las preguntas de la Prueba Saber de Matemáticas en los contenidos de los estándares?

Antes de presentar más información sobre el estudio, es importante aclarar que aunque el término “alineación” se utiliza algunas veces en el área de evaluación y en el ámbito de la psicometría para referirse a varias cosas. Por ejemplo, alineación también puede referirse a equiparación de puntajes entre pruebas o la relación entre pruebas y diferentes aspectos del proceso educativo, incluyendo estándares académicos, el proceso de enseñanza y aprendizaje, y al proceso de evaluación en el aula de clase. En este estudio, el concepto de alineación se utiliza para referirse a la alineación entre una prueba académica y una serie de estándares académicos únicamente y se considera parte de la validez de la prueba.

De igual manera, es importante aclarar que la alineación que se presenta en este estudio se centra únicamente en la alineación entre prueba y estándares. Es importante recordar que también es de suma importancia examinar la alineación entre los estándares académicos y las prácticas curriculares (Matore y Sireci, 2009), entre las pruebas académicas y las prácticas curriculares (McGehee y Griffith, 2001), y entre las pruebas censales y las prácticas de evaluación o la forma como se evalúa en el aula de clase (Ruíz-Primo et al, 2012). También es interesante examinar la forma como se implementan los estándares en el aula de clase y la forma cómo lo utilizan los profesores. Esperamos que este estudio de las bases para que se hagan otros estudios examinando otros tipos de alineación y las formas como los estándares se implementan en clase. Todos estos tipos de estudio permiten tener un mejor entendimiento sobre la alineación ya que la interpretación final de las puntuaciones de una prueba dependerá también de los aspectos anteriormente mencionados.

Los Estándares Básicos de Competencias en Matemáticas

Los Estándares Básicos de Competencias en Matemáticas fueron desarrollados por el Ministerio de Educación Nacional en colaboración con la Asociación Colombiana de Facultades de Educación –Ascofade–, profesores de

Matemáticas y miembros de la comunidad educativa. Estos estándares describen las competencias y habilidades necesarias que los estudiantes deben tener al finalizar cada año escolar y tienen como finalidad mejorar el nivel de la educación. Los Estándares Básicos de Competencias en Matemáticas están divididos en cinco ejes básicos: 1) pensamiento numérico y sistemas numéricos, 2) pensamiento espacial y sistemas geométricos, 3) pensamiento métrico y sistemas de medida, 4) pensamiento aleatorio y sistemas de datos, y 5) pensamiento variacional y sistemas algebraicos y analíticos, y cada uno de ellos tiene varios logros (ver Anexo 1).

Las Pruebas Saber de Matemáticas

Las Pruebas Saber son una batería de exámenes en las áreas de Lenguaje, Matemáticas, Ciencias Sociales y Ciencias Naturales. Estas son unas pruebas censales que toman los estudiantes en los grados 5º y 9º. Según los diseñadores de las Pruebas Saber, el objetivo de éstas son las de “obtener, procesar, interpretar y divulgar información confiable de tal manera que se constituyan en una base sólida para la toma de decisiones en las diferentes instancias del servicio educativo” (ICFES, 2008). Para este estudio se utilizó la prueba para grado 9º del año 2009, la cual tienen quince preguntas de selección múltiple con cuatro opciones (ver más información en <http://www.icfes.gov.co>). Vale aclarar que la alineación de la prueba con los estándares depende de la forma cómo se administra la prueba y de quién la administra. Como el grado de control de la administración condiciona la calidad de la información, es importante que la prueba se administre consistentemente para garantizar que la información que se recoja sea válida.

Marco conceptual

El tipo de alineación que se examina en este estudio también se conoce con el nombre de alineación o articulación horizontal entre el contenido de los estándares y lo que se está midiendo en una evaluación externa (Martineau et al., 2007). Algunos autores consideran que el grado de alineación entre una evalua-

ción y unos estándares académicos es una condición necesaria para validar exámenes que evalúan el aprendizaje (Messick, 1989; Webb, 1997). En este estudio, el grado de alineación se evaluó usando el modelo propuesto por Webb (1997). Algunos estudios han mostrado que la metodología de Webb es la más adecuada para determinar el grado de alineación entre las pruebas censales y los estándares académicos ya que examina la alineación desde varias perspectivas (Herman, Webb, y Zuniga, 2007; Martone y Griffith, 2009). En este modelo, el grado de alineación se mide con base en cuatro criterios: 1) en el contenido que se está evaluando, 2) en la demanda cognitiva de las preguntas en las evaluaciones, 3) en el número de contenidos de los estándares que se están evaluando, y 4) en la forma como están distribuidas las preguntas de las evaluaciones en los estándares.

Alineación con base en contenido

Una condición importante que se debe considerar en los estudios de alineación es si el contenido que se evalúa en un examen está relacionado con los contenidos descritos en los estándares (Webb, 1997; Webb, Herman & Webb, 2007). Por contenido se entienden los temas, conocimientos, habilidades o competencias que se están evaluando (AERA, APA, NCME, 1999).

Alineación con base en demanda cognitiva

Para evaluar el grado de alineación entre unos estándares y un examen también se analiza el nivel de demanda cognitiva de los contenidos de los estándares y el nivel de demanda cognitiva de las preguntas de la evaluación (Webb, 1997; Webb et al., 2007). A continuación se describen los niveles de demanda cognitiva usados por Webb (1997).

Nivel 1 – Recordar y reproducir. En el nivel 1 el estudiante solamente tiene que identificar o recordar la respuesta y no tiene la necesidad de hacer ningún tipo de razonamiento para contestar una pregunta. Algunos ejemplos incluyen: medir un ángulo; encontrar el área de un rectángulo; multipli-

car dos números; convertir un número en notación científica a número decimal; identificar una diagonal en una figura geométrica.

Nivel 2 – Habilidades y conceptos. En el nivel 2 el estudiante tiene que realizar algún tipo de proceso mental que va más allá de simplemente recordar o reproducir una respuesta, por lo tanto este proceso es más complejo que en el nivel 1. Estos procesos requieren que el estudiante tome una decisión sobre cómo contestar una pregunta o cómo resolver un problema. Algunos ejemplos incluyen: extender un patrón geométrico; clasificar cuadriláteros; organizar un conjunto de datos y construir una tabla; determinar una estrategia para estimar el número de monedas en un recipiente; comparar dos conjuntos de datos usando la media, mediana y moda de cada conjunto.

Nivel 3 – Pensamiento estratégico. En el nivel 3 se requiere que el estudiante use procesos cognitivos más complejos que en los dos primeros niveles. Por ejemplo, a este nivel el estudiante tiene que razonar, planear o usar evidencias. En general, se requiere que se contesten las preguntas usando varios pasos, que justifiquen las respuestas, que saquen conclusiones con base en observaciones y que expliquen fenómenos entre otros. Algunos ejemplos incluyen: explicar cómo los cambios en las dimensiones de una figura geométrica afectan su área o perímetro; dar una justificación matemática para resolver un problema; interpretar información de una figura; escribir una regla matemáticas para un patrón no rutinario; plantear y resolver ecuaciones para un problema dado.

Nivel 4 – Pensamiento extendido. En el nivel 4 se requiere el uso de procesos cognitivos complejos. Se tienen que hacer conexiones entre ideas y conceptos, escoger entre varias alternativas para resolver un problema, o aplicar los resultados de un experimento en otros contextos. Muchas de estas actividades requieren que se hagan sobre un período extendido de tiempo, pero el

hecho que una actividad tome bastante tiempo no es requisito para ser de este nivel. Algunos ejemplos incluyen: recolectar datos teniendo en cuenta varias variables y analizar los resultados; plantear una regla para un patrón complejo y encontrar un fenómeno que exhiba ese comportamiento; modelar un fenómeno social con varias alternativas y escoger un método para resolverlo usando un modelo matemático.

Alineación con base en cobertura

Este criterio se usa para determinar si los contenidos que abarcan las preguntas de la evaluación cubren los contenidos que se abarcan en los estándares (Webb, 1997). Bajo este criterio solamente se tiene en cuenta el número de contenidos de los estándares que están siendo evaluados por lo menos por una pregunta. En este criterio no se tiene en cuenta el número de veces que está siendo evaluado un determinado contenido en los estándares.

Alineación con base en énfasis

Este criterio se usa para indicar qué tanto énfasis se da a unos contenidos que a otros en una evaluación (Webb, 1997). A diferencia del criterio anterior, acá sí se tiene en cuenta el número de veces que está siendo evaluado un determinado contenido. Por lo tanto, se tiene en cuenta la diferencia en la proporción de los contenidos en cada eje temático y la proporción de preguntas relacionadas con cada uno de estos.

Estudios previos han indicado que existen varios grados de alineación (Bhola, Impara y Buchendahl, 2003). Herman, Webb y Zuniga (2007) encontraron que diferentes grupos de jueces pueden producir diferentes resultados; por lo tanto es de suma importancia entrenar adecuadamente a los jueces para minimizar este efecto. Los jurados deben familiarizarse con la prueba, los estándares y el proceso de codificación. En cuanto a la prueba, los jueces deben conocer el formato, el tipo de preguntas, los recursos disponibles para tomarla (por ejemplo, calculadores, diccionarios, etc.) y la

forma como se administra (Bhola, Impara y Buchendahl, 2003). En cuanto a los estándares, los jueces deben conocer cómo están organizados, cómo se utilizan en el aula de clase y cómo deben evaluarse (López, 2009).

Igualmente, los resultados aportan información importante para mejorar las pruebas (Lopez, Webb y Stansfield, 2006). La alineación da información sobre qué tipo de inferencias se pueden hacer sobre el proceso de aprendizaje sobre todo si se van a tomar decisiones importantes con base en los resultados de las mismas (Pehuniak, 2005). Por otro lado, estudios previos han encontrado que las preguntas de construcción (por ejemplo, las preguntas abiertas, los ensayos, etc.) tienden a estar alineadas con varios estándares ya que generalmente requieren que los estudiantes apliquen diferentes tipos de conocimientos o habilidades para contestarlas (Webb, 2002).

Metodología

En esta sección se reportan los resultados de un estudio cuantitativo no experimental, que se complementan con datos cualitativos sobre la percepción de los jueces acerca de las preguntas y el nivel de alineación entre la evaluación externa y los estándares. Los datos para este estudio se recolectaron en una sesión de cinco horas.

Participantes

Este estudio contó con la participación de siete jueces. En la Tabla 1 aparece información sobre cada uno de ellos. Su tarea era la de codificar los logros de los Estándares Básicos de Competencias en Matemáticas en grado 9º con base en el nivel de demanda cognitiva. Adicionalmente, los jueces codificaron las preguntas de las Pruebas Saber de Matemáticas del 2006 con base en el nivel de demanda cognitiva y los contenidos en los estándares. Finalmente, cada juez hizo comentarios sobre aspectos problemáticos con las preguntas y acerca de sus percepciones sobre el grado de alineación entre la evaluación externa y los estándares.

Tabla 1. Información sobre los jueces

Juez	Formación	Experiencia docente
1	Licenciada en Matemáticas; candidata a Magíster en educación	Profesora de Matemáticas en educación básica y media (2 años); profesora de Matemáticas en educación superior (3 años)
2	Licenciada en Matemáticas	Profesora de Matemáticas en educación básica y media (6 años); profesora de Matemáticas en educación superior (4 años)
3	Licenciado en Matemáticas; Magíster en Matemáticas	Profesor de Matemáticas en educación básica y media (2 años); profesor de Matemáticas en educación superior (3 años)
4	Licenciada en Matemáticas; candidata a Magíster en Matemáticas	Profesora de Matemáticas en educación básica y media (3 años); profesora de Matemáticas en educación superior (3 años)
5	Licenciada en Matemáticas; candidata a Magíster en Matemáticas	Profesora de Matemáticas en educación básica y media (5 años)
6	Licenciado en Matemáticas	Profesor de Matemáticas en educación básica y media (2 años); diseñador de software educativo para Matemáticas (1 año)
7	Licenciada en Matemáticas	Profesora de Matemáticas en educación básica y media (5 años); profesora de Matemáticas en educación superior (1 año)

Recolección de datos

Este estudio se realizó en cuatro fases para facilitar el proceso de evaluación del grado de alineación entre los Estándares Básicos de Competencias en Matemáticas y la Prueba Saber en grado 9°.

Fase uno. En la primera fase se hizo una sesión de entrenamiento a los jueces para familiarizarlos con el proceso de alineación, con los niveles de demanda cognitiva que se usaron para codificar los estándares y las pruebas, con los Estándares Básicos de Competencias en Matemáticas y con la Prueba Saber de Matemáticas en grado 9°. Se hicieron unos ejercicios para que los jueces pudieran practicar cada uno de estos aspectos y para que pudieran aclarar dudas.

Fase dos. Después del entrenamiento, cada juez codificó cada estándar. Luego todos los jueces se reunieron con los investigadores para compartir y justificar sus codificaciones. Al final se llegó a un consenso sobre el nivel de demanda cognitiva de cada uno de los contenidos en los estándares (ver Anexo 1).

Fase tres. En la tercera fase del estudio cada juez le asignó un nivel de demanda cognitiva y hasta tres contenidos a cada pregunta en la prueba usando un formato de codificación (ver Anexo 2). En esta fase, los jueces también escribieron comentarios sobre cada pregunta en el formato de codificación.

Fase cuatro. En la última fase se reunieron los jueces con el investigador para hablar sobre sus impresiones sobre el grado de alineación entre la Prueba Saber de Matemáticas en grado 9° y los Estándares Básicos de Competencias en Matemáticas.

Análisis de datos

Los datos cuantitativos se analizaron usando cuatro criterios diferentes para establecer el grado de alineación entre los estándares académicos y la Prueba Saber de Matemáticas en grado 9°.

Alineación con base en el contenido. Según Webb (1997) este criterio se evaluó teniendo en cuenta el número de preguntas en el examen que están relacionadas con los logros en los estándares. Para que haya un grado aceptable de alineación a nivel de contenido, se espera que por lo menos seis preguntas estén evaluando los contenidos descritos en cada eje básico. Este número se toma con base en un estimado del número de preguntas que pueden producir una subescala confiable para hacer inferencias sobre el grado de competencia o habilidad que tiene el alumno sobre ese contenido (Subkoviak, 1998). Si hay seis o más preguntas relacionadas, se considera que el grado de alineación es fuerte. Si hay cinco preguntas relacionadas, se considera que el grado de alineación es débil y si hay menos de cinco preguntas asociadas con un eje temático, se considera que el grado de alineación no es adecuado.

Alineación con base en la demanda cognitiva. Para analizar este criterio se comparó el nivel de demanda cognitiva de cada pregunta con el nivel de demanda cognitiva de los contenidos en cada eje temático que están relacionadas con cada una. Para que haya un grado de alineación adecuado, el 50 por ciento de las preguntas tienen que tener por lo menos el mismo nivel de demanda cognitiva. Si el porcentaje de preguntas que está por lo menos al mismo nivel está entre 40-50 por ciento, se considera que el nivel de alineación es débil y si el porcentaje de preguntas que tiene por lo menos el mismo nivel de demanda cognitiva es inferior al 40 por ciento, se considera que el nivel de alineación entre las preguntas y los ejes básicos no es adecuado.

Alineación con base en la cobertura. Para analizar qué tanta cobertura tienen las preguntas de las pruebas en cada eje temático se tomó en cuenta el número de logros que tienen por lo menos una pregunta relacionada. Si por lo menos el 50 por ciento de los contenidos en un eje temático tiene por lo menos una pregunta relacionada en la prueba se considera que el nivel de alineación es adecuado. Si entre el 40-50 por ciento de los contenidos en un eje temático tiene por lo menos una pregunta relacionada, se considera que el nivel de alineación es débil y si el porcentaje es inferior al 40 por ciento, se considera que el nivel de alineación entre las preguntas y el eje temático no es adecuado.

Alineación con base en el énfasis. Para analizar cómo están distribuidos los contenidos de los estándares en la Prueba Saber de Matemáticas en grado 9º, se calculó un índice. Este índice tiene en cuenta el número de logros en cada eje temático que están relacionados al menos con una de las preguntas. El índice se calcula teniendo en cuenta la diferencia en la proporción de los contenidos y la proporción de preguntas relacionadas con cada uno de estos. Si el índice equivale a 1 esto significa que hay una distribución perfecta y asume que el número de preguntas relacionadas a un eje básico están distribuidas equitativamente entre

los contenidos de cada eje temático. Por el contrario, si el índice está cerca a 0 esto significa que solamente unos pocos contenidos están siendo evaluados por las preguntas de la prueba. Por lo tanto, si el índice es superior a .7, se considera que el nivel de alineación es adecuado. Si el índice es entre .6 y .7, se considera que el nivel de alineación es débil y si el índice es inferior a .6, se considera que el nivel de alineación entre las preguntas y el eje temático no es adecuado.

Comentarios y percepciones. Los comentarios y las percepciones de los jueces se analizaron cualitativamente utilizando el método interpretativo de Hatch (2002). Se transcribieron todos los comentarios de los jueces y se archivaron en documentos de Word. Luego se leyeron todos los comentarios y percepciones para establecer puntos comunes y discrepancias que permitieran hacer generalizaciones sobre aspectos problemáticos.

Para darle mayor credibilidad y validez a este estudio se entrenó a los jueces en el proceso de codificación. Para determinar la concordancia entre los jueces, se calculó el coeficiente alfa de Cronbach (Cronbach, 1971). En este estudio se determinó que la correlación interna en la codificación entre los siete jueces fue de .92 (ver Tabla 2). Estas correlaciones son aceptables y sugieren que hubo bastante acuerdo entre los jueces al asignar los niveles de demanda cognitiva entre las preguntas y los logros. El grado de relación entre pares de jueces al asignar logros a las preguntas fue de .61 y la correlación interna entre los siete jueces fue de .92. Esto indica que el entrenamiento que se les hizo a los jueces fue bueno y por lo tanto se tiene mucha confianza a la codificación que hicieron. Webb et al. (2007) resaltan la importancia de entrenar adecuadamente a los jueces en los estudios de alineación.

Tabla 2. Confiabilidad entre jueces

Grado	Correlación interna entre jueces	Comparación entre pares	Número de preguntas	Número de jueces
9	.92	.61	15	7

Resultados

Grado de alineación a nivel de contenido

Se encontró que la Prueba de Saber de Matemáticas en grado 9° no está alineada fuertemente con ninguno de los cinco ejes temáticos (ver Tabla 3). Con base en estos resultados se

puede concluir que la Prueba Saber de Matemáticas en grado 9° no está alineada a nivel de contenido. Esto indica que el número de preguntas relacionadas con cada uno de los cinco ejes temáticos no es el más adecuado.

Tabla 3. Alineación con base en el contenido

Ejes temáticos	N° de Contenidos	Contenidos relacionados	Alineación a nivel de contenido
Nombre		N° de preguntas	
1 – Pensamiento numérico y sistemas numéricos	4	5	Débil
2 - Pensamiento espacial y sistemas geométricos	4	1	Inadecuada
3 - Pensamiento métrico y sistemas de medidas	3	1	Inadecuada
4 - Pensamiento aleatorio y sistemas de datos	9	4	Inadecuada
5 - Pensamiento variacional y sistemas algebraicos y analíticos	9	5	Débil

Grado de alineación a nivel de demanda cognitiva

Se encontró que la Prueba Saber de Matemáticas en grado 9° tiene un grado de alineación parcialmente adecuado (ver Tabla 4). El porcentaje de preguntas relacionadas en cuatro de los cinco de los ejes básicos están por lo menos al mismo nivel de demanda cognitiva. Cabe anotar que en uno de estos ejes temáticos, ‘Pensamiento espacial y sistemas geométricos’, solamente tiene una pregunta relacionada

y por lo tanto es difícil hacer inferencias válidas sobre el grado de alineación. El único eje temático que no está alienado adecuadamente a nivel de demanda cognitiva es ‘Pensamiento métrico y sistemas de medidas’. En este eje temático solamente el 20 por ciento de las preguntas relacionadas están por lo menos al mismo nivel de demanda cognitiva que los objetivos.

Tabla 4. Alineación con base en la demanda cognitiva

Ejes temáticos	N° de contenidos	N° de preguntas relacionadas	Nivel de las preguntas con relación a los estándares			Alineación a nivel de demanda cognitiva
			Promedio Menor	Promedio Igual	Promedio Mayor	
1 – Pensamiento numérico y sistemas numéricos	4	5	11	50	39	Adecuado
2 - Pensamiento espacial y sistemas geométricos	4	1	0	71	29	Adecuado
3 - Pensamiento métrico y sistemas de medidas	3	1	80	10	10	Inadecuada
4 - Pensamiento aleatorio y sistemas de datos	9	4	38	45	17	Adecuada
5 - Pensamiento variacional y sistemas algebraicos y analíticos	9	5	20	12	68	Adecuada

Grado de alineación a nivel de cobertura

En la Prueba de Matemáticas de Grado 9°, solamente uno de los ejes temáticos, ‘Pensamiento numérico y sistemas numéricos’ está adecuadamente alineado a nivel de cobertura

(Ver Tabla 5). En este eje, el 57 por ciento de los contenidos tienen por lo menos una pregunta relacionada. En los otros cuatro ejes temáticos encontramos que estos no tienen por lo menos el 50 por ciento de los contenidos

con una pregunta relacionada. Consecuentemente, se puede concluir que la cobertura no es muy buena. Esto indica que muchos de los

contenidos de los estándares no se están evaluando.

Tabla 5. Alineación con base en el cubrimiento

Ejes temáticos	Nº Contenidos	Nº de preguntas relacionadas	Cobertura de objetivos		Alineación a nivel de cobertura
			Nº de contenidos relacionados	Porcentaje del total	
Nombre			Promedio	Promedio	
1 – Pensamiento numérico y sistemas numéricos	4	5	2.29	57	Adecuada
2 - Pensamiento espacial y sistemas geométricos	4	1	1	25	Inadecuada
3 - Pensamiento métrico y sistemas de medidas	3	1	0.71	24	Inadecuada
4 - Pensamiento aleatorio y sistemas de datos	9	4	2.71	30	Inadecuada
5 - Pensamiento variacional y sistemas algebraicos y analíticos	9	5	3.57	40	Débil

Grado de alineación a nivel de distribución

Por último, también se encontró que el grado de alineación con base en énfasis en grado 9º no es adecuado en cada uno de los cinco ejes básicos (ver Tabla 6). Uno de los ejes temáticos ‘Pensamiento métrico y sistemas de medidas’ no cumplió con los requisitos. El índice

para este eje temático es de solamente .55. Otro eje temático, ‘Pensamiento numérico y sistemas numéricos’ tiene un índice de .66. Por lo tanto se concluye que el nivel de alineación es débil. Con base en esto índices, se infiere que no se le está dando mucho énfasis a estos dos ejes temáticos.

Tabla 6. Alineación con base en el énfasis – Saber Ciencias Matemáticas 9º

Ejes temáticos		Índice de énfasis		Alineación a nivel de énfasis
		% preguntas relacionadas con el eje temático/Total relacionadas	Índice	
Nombre	Nº Contenidos	Prom.	Prom.	
1 - Pensamiento numérico y sistemas numéricos	4	31	0.66	Débil
2 - Pensamiento espacial y sistemas geométricos	4	6	0.86	Adecuada
3 - Pensamiento métrico y sistemas de medidas	3	5	0.55	Inadecuada
4 - Pensamiento aleatorio y sistemas de datos	9	26	0.85	Adecuada
5 - Pensamiento variacional y sistemas algebraicos y analíticos	9	32	0.86	Adecuada

Comentarios de los jueces

Los jueces hicieron comentarios sobre ocho de las quince preguntas. En general, todos los comentarios tienen que ver con la falta de alineación de estas preguntas con los contenidos descritos en los estándares.

Percepciones

Es importante resaltar que los resultados descritos en esta sección solamente indican las percepciones de los jueces y no representan un análisis riguroso del grado de alineación entre los Estándares Básicos de Competencias en

Matemáticas y la Prueba Saber de Matemáticas en grado 9°. Los siete jueces consideraron que la prueba evalúa los contenidos más importantes descritos en los estándares, pero los contenidos no tienen suficientes preguntas para poder tomar decisiones importantes sobre las competencias de los estudiantes. Igualmente, los jueces creen que el nivel de demanda cognitiva de la prueba es comparable con el nivel cognitivo de los contenidos en los estándares. Finalmente, los jueces expresaron que la prueba Saber de Matemáticas en grado 9° y los Estándares Básicos de Competencias en Matemáticas están parcialmente alineados, pero el nivel de alineación entre ellos se puede mejorar.

Discusión

Uno de los aspectos más importantes del modelo Webb (1997) es que la alineación entre las evaluaciones externas y los estándares se analiza desde varias perspectivas. En términos generales, se encontró que el grado de alineación no es el más adecuado. A nivel de contenido, se evidencia que el grado de alineación no es muy fuerte, pero esto es un resultado esperado ya que el número de preguntas en la prueba Saber de Matemáticas en grado 9° es muy bajo. Esto indica que se debe tener mucho cuidado con la forma cómo se interpretan los resultados, sobre todo en los ejes básicos donde no hay un número adecuado de preguntas en la prueba. La articulación entre el contenido de las pruebas y el contenido de los estándares se puede corregir aumentando el número de preguntas pero hay que tener en cuenta que los estudiantes están tomando otras pruebas el mismo día. Por lo tanto, es importante examinar cuántas preguntas se pueden agregar sin que esto vaya a causar algún estrés o cansancio en los estudiantes. También existe la posibilidad que no se puedan agregar preguntas a la prueba. Si este es el caso, habría que mirar cómo se pueden usar preguntas integradas para que estén alineadas a más de un eje curricular.

Por otro lado, a nivel de demanda cognitiva, se encontró que el grado de alineación es adecuado. Estos resultados sugieren que el nivel

de exigencia de las pruebas está alineado con las expectativas que se describen en los contenidos de los ejes temáticos de los estándares y que en teoría la prueba Saber de Matemáticas en grado 9° no es ni muy difícil, ni muy fácil. Lo más importante es tener una evaluación que mida diferentes tipos de habilidades o competencias. Lo único preocupante es que la prueba Saber de Matemáticas en grado 9° solamente utiliza preguntas de selección múltiple. Como es de conocimiento general, este tipo de pregunta tiene muchas limitaciones, sobre todo si lo que se quiere es medir competencias de alta complejidad cognitiva (AERA, APA, NCME, 1999). Como se tiene esta limitación, muchos de los contenidos en los estándares solamente se pueden evaluar en el aula de clase. Habría que examinar cómo se están evaluando estas competencias en el aula de clase.

También se encontró que hay muchos contenidos que no se están evaluando en la prueba Saber de Matemáticas en grado 9°, pero como se mencionó anteriormente, esto es de esperar ya que el número de preguntas es muy bajo y no permite una representatividad adecuada de los contenidos en los estándares. Esto indica que las preguntas en la prueba Saber de Matemáticas en grado 9° no son una muestra representativa de todos los contenidos y las competencias descritas en los cinco ejes básicos. Por ende se debe tener mucha cautela al interpretar los resultados en las mismas. Esto también implica que los demás contenidos descritos en los estándares se deben evaluar en el aula de clase antes de tomar cualquier tipo de decisión que tenga que ver con el proceso de enseñanza y aprendizaje. Especialmente si existe la tendencia de solamente enfocar la enseñanza en lo que se mide en las evaluaciones, corriendo el riesgo que muchos docentes se enfoquen en algunos contenidos y competencias de los estándares e ignoren otros (Herman, 2004). Esto es una consecuencia o impacto de las pruebas, así no sea intencional, que se debe evitar a toda costa (Herman, 2004; López, 2008; Messick, 1989).

En cuanto al énfasis que se le da a los contenidos que están siendo evaluados, se encontró que las preguntas en la prueba Saber de Matemáticas en grado 9° están representadas proporcionalmente en estos contenidos. Esto es un resultado importante, pero con el bajo número de preguntas que se tiene, esto es relativamente fácil de alcanzar. Es fundamental recordar que si el número de preguntas aumenta o si la cobertura de los logros es mayor, la distribución de los contenidos evaluados tiene que ser proporcional, ya que una vez las evaluaciones se vuelvan de conocimiento público, éstas pueden indicar o sugerir qué se debe estar enseñando en el aula y qué deben estar aprendiendo los estudiantes. Existe la posibilidad que inconscientemente se le esté dando mayor prioridad a ciertos contenidos o competencias que a otros.

Aunque no existe una fórmula que indique cuál debe ser el grado de alineación ideal entre las evaluaciones externas y unos estándares, se espera que éste sea lo suficientemente fuerte para que se puedan tomar decisiones acertadas sobre el proceso de enseñanza y aprendizaje. De ninguna manera se espera que haya un grado de alineación perfecto, ya que esto es casi imposible de lograr por diferentes factores. En este estudio, se encontró que hay varios de ellos que no permiten que el grado de alineación entre la prueba Saber de Matemáticas en grado 9° y los Estándares no sea mayor.

El primer factor, tal como se había mencionado anteriormente, es que el número de preguntas en la prueba Saber de Matemáticas en grado 9°, al igual que en muchas evaluaciones estandarizadas, es relativamente muy bajo, lo cual no permite que se pueda evaluar una muestra representativa de los contenidos en cada uno de los cinco ejes temáticos. De igual manera, el tipo de pregunta o formato que se usa, de selección múltiple, limita el número de contenidos que se puedan evaluar ya que muchos de estos no se pueden medir de esta manera. Por ejemplo, algunos contenidos requieren que los estudiantes produzcan resultados con base en trabajo extensos –proyectos– y

éstos son muy difíciles de medir en una evaluación estandarizada externa. Por lo tanto algunos contenidos están diseñados para ser evaluados solamente dentro del aula de clase.

El nivel de especificidad de los estándares también podría estar afectando la articulación entre los contenidos y las preguntas en la prueba Saber de Matemáticas en grado 9°. Así, los jueces comentaron que algunos contenidos son muy globales y otros son muy específicos. Algunos estudios han mostrado que entre menos específicos sean los estándares, mayor tiende a ser el grado de alienación (López, Webb & Stansfield, 2006). Por otro lado, falta mayor claridad en algunos contenidos y muchos de estos tienden a repetirse o son muy parecidos. Esto hace que sea difícil escribir preguntas que los evalúen y por lo tanto el grado de alineación entre estas preguntas y los estándares se hace menos evidente. Para finalizar, también se encontró que como no hay pruebas Saber de Matemáticas en todos los grados, existe la posibilidad que una pregunta esté midiendo contenidos que están descritos en grados inferiores y no necesariamente en los estándares diseñados para el grado que se está evaluando.

Con base en los comentarios de los jueces también se puede concluir que es importante revisar los estándares para verificar que los contenidos sean claros y evaluables. El bajo grado de alineación no sólo puede indicar que puede haber problemas con las evaluaciones sino que también puede haber problemas con los estándares. Por eso es importante recordar que durante el proceso de diseño de los estándares hay que pensar en la forma en que se pueden medir los logros que se proponen. De igual manera es importante que los estándares también tengan una sección donde se describa cómo se deben evaluar cada uno de los contenidos. Este proceso podría no sólo facilitar el diseño de sistemas de evaluaciones estandarizadas externas –las Prueba Saber– sino también las evaluaciones diseñadas por los mismos profesores y que se puedan usar en el aula de clase.

Comentarios finales

Los resultados de este estudio tienen varias implicaciones. Primero, los diseñadores de las evaluaciones estandarizadas podrían tener en cuenta los resultados de los estudios de alineación para hacer cambios en la evaluación y para identificar áreas o contenidos que no están siendo evaluados adecuadamente. Lo más importante es que los diseñadores tengan siempre presente los estándares en el diseño de las preguntas. Por ejemplo, las especificaciones de las evaluaciones deben establecer claramente la conexión entre cada pregunta y los contenidos descritos en los estándares. Estas especificaciones también deberían dar información sobre el contenido, la demanda cognitiva, el tipo de pregunta, la forma como se debe calificar y la forma como se deben interpretar los resultados. Igualmente, los diseñadores de las evaluaciones estandarizadas también podrían usar los resultados de un estudio de alineación como evidencia de la validez de contenido de la evaluación.

Segundo, se evidencia la necesidad de complementar los resultados de las evaluaciones estandarizadas con los resultados de las evaluaciones que se desarrollan en el aula de clase, ya que los resultados de las evaluaciones estandarizadas no muestran necesariamente todo lo que los estudiantes han tenido la oportunidad de aprender del currículo que se está enseñando en el aula de clase. Por lo tanto es muy importante que se siga trabajando con los docentes para que se familiaricen con los estándares y para que puedan usarlos y evaluarlos adecuadamente en el aula de clase. Otro aspecto importante en el proceso de enseñanza y aprendizaje es entrenar a los docentes a interpretar los resultados de las evaluaciones estandarizadas para que puedan tomar decisiones informadas que mejoren el nivel de la calidad de la educación.

Por último, hay que recalcar la importancia de interpretar adecuadamente los resultados de las evaluaciones estandarizadas. Como generalmente la totalidad de los contenidos de los estándares no se pueden evaluar todos en una

sola evaluación, es importante resaltar que no se pueden tomar decisiones relevantes sobre el proceso de enseñanza y aprendizaje únicamente con base en los resultados de las evaluaciones estandarizadas. Es necesario utilizar varios criterios, incluyendo las evaluaciones que están haciendo los docentes en el aula de clase, para determinar qué y cuánto están aprendiendo los estudiantes. Estudios como el que presentó en este documento dan claridad sobre cómo evaluar la alineación de una prueba y plantean la necesidad de hacer estudios de este tipo, basados en juicio de expertos, para abordar aspectos básicos de la validez de constructo y de contenido de las pruebas.

Referencias

- AERA, APA & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bhola, D. S., Impara, J. C., & Buchendahl, C. W. (2003). Aligning tests with states content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21-29.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). *Educational Measurement (2nd Ed.)*. Washington, D. C.: American Council on Education.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: SUNNY Press.
- Herman, J. (2004). The effects of testing in instruction. En S. Fuhrman and R. Elmore (Eds.), *Redesigning accountability systems for education (pp. 141-166)*. New York: Teachers College Press.
- Herman, J. L., Webb, N. M. & Zuniga, S. A. (2007). Measurement issues in the alignment of standards and assessments: A case study. *Applied Measurement in Education*, 20(1), 101-176.
- ICFES (2008). *Documentos y Marco Legal pruebas SABER*. Bajado el 7 de agosto de 2008, de <http://www.icfes.gov.co>
- Lopez, A. A. (2009). Estudio de alineación de las pruebas Saber de Ciencias. En J. Montoya (Eda.), *Educación para el siglo XXI: Aportes*

- del Centro de Investigación y Formación en Educación, CIFE, 2001-2008 (pp. 551-582). Bogotá: Ediciones Uniandes.
- López, A. (2008). *Potential impact of language tests: Examining the alignment between testing and instruction*. Saarbrücken, Germany: VDM Publishing.
- López, A., Webb, N., y Stansfield, C. (2006). *Alignment the New Mexico Language Arts Frameworks and the Spanish Reading Standards-Based Assessments*. Report submitted to the New Mexico Public Education Department. Rockville, MD: Second Language Testing, Inc.
- Martone, A. & Sireci (2009). Evaluating alignment between curriculum, assessment and instruction. *Review of Educational Research*, 79(3), 1-76.
- Martineau, J., Paek, P., Keene, J. & Hirsch, T. (2007). Integrated comprehensive alignment as a foundation for measuring student progress. *Educational Measurement: Issues and Practice*, 26(2), 28-35.
- McGehee, J. J., & Griffith, L. K. (2001). Large-scale assessments combined with curriculum alignment: Agents of change. *Theory Into Practice*, 40(2), 137-144.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: Macmillan.
- Pehuniak, P. (2004). Educational assessment in an era of accountability. En J. E. Wall & G. R. Walz (Eds.). *Measuring up: Assessment issues for teachers, counselors, and administrators*. Greensboro, NC: CAPS Press.
- Porter, A. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31, 3-14.
- Rothman, R. (2004). Benchmarking and alignment of state standards and assessment. En S. Fuhrman and R. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 96-114). New York: Teachers College Press.
- Ruíz-Primo, M. A., Li, M., Wills, K., Giamellaro, M., Lan, M. C., Mason, H., & Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. *Journal of Research in Science Teaching*, 49(6), 691-712.
- Subkoviak, M. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Webb, N. (2002). *An analysis of the alignment between standards and assessments for three states*. Madison, WI: Wisconsin Center for Education Research.
- Webb, N. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education. Madison, Wisconsin: Wisconsin Center for Education Research, University of Wisconsin.
- Webb, N., Herman, J. & Webb, N. (2007). Alignment of mathematics state-level standards and assessment: The role of reviewer agreement. *Educational Measurement: Issues and Practice*, 26(2), 17-29.
- Wise, L. (2004). Vertically-articulated content standards. Disponible en http://www.nciea.org/publications/RILS_LW04.pdf.

ANEXOS

Anexo 1

Consenso sobre la demanda cognitiva de los logros en los estándares
Grado 9

Nivel	Descripción	NDC
1	Pensamiento numérico y sistemas numéricos	2
1.1	Utilizo números reales en sus diferentes representaciones y en diversos contextos	1
1.2	Resuelvo problemas y simplifico cálculos usando propiedades y relaciones de los números reales y de las relaciones y operaciones entre ellos	2
1.3	Utilizo la notación científica para representar medidas de cantidades de diferentes magnitudes	1
1.4	Identifico y utilizo la potenciación, la radicación y la logaritmación para representar situaciones Matemáticas y no Matemáticas y para resolver problemas	2
2	Pensamiento espacial y sistemas geométricos	3
2.1	Conjeturo y verifico propiedades de congruencias y semejanzas entre figuras bidimensionales y entre objetos tridimensionales en la solución de problemas	3
2.2	Reconozco y contrasto propiedades y relaciones geométricas utilizadas en demostración de teoremas básicos (Pitágoras y Tales)	2
2.3	Aplico y justifico criterios de congruencias y semejanzas entre triángulos en la resolución y formulación de problemas	3
2.4	Uso representaciones geométricas para resolver y formular problemas en las Matemáticas y en otras disciplinas	3
3	Pensamiento métrico y sistemas de medidas	2
3.1	Generalizo procedimientos de calculo validos para encontrar el área de regiones planas y el volumen de sólidos	3
3.2	Selecciono y uso técnicas e instrumentos para medir longitudes, áreas de superficies, volúmenes y ángulos con niveles de precisión apropiados	2
3.3	Justifico la pertinencia de utilizar unidades de medida estandarizadas en situaciones tomadas de distintas ciencias	2
4	Pensamiento aleatorio y sistemas de datos	2
4.1	Reconozco como diferentes maneras de presentación de información pueden originar distintas interpretaciones	2
4.2	Interpreto analítica y críticamente información estadística proveniente de diversas fuentes (prensa, revistas, televisión, experimentos, consultas, entrevistas)	3
4.3	Interpreto y utilizo conceptos de media, mediana y moda	3
4.4	Selecciono y uso algunos métodos estadísticos adecuados al tipo de problema, de información y al nivel de la escala en la que esta se representa (nominal, ordinal, de intervalo o de razón)	2
4.5	Comparo resultados de experimentos aleatorios con los resultados previstos por un modelo matemático probabilística	2
4.6	Resuelvo y formulo problemas seleccionando información relevante en conjuntos de datos provenientes de fuentes diversas (prensa, revistas, televisión, experimentos, consultas, entrevistas)	3
4.7	Reconozco tendencias que se presentan en conjuntos de variables relacionadas	1
4.8	Calculo probabilidad de eventos simples usando métodos diversos (listado, diagramas de árbol, técnicas de conteo)	2
4.9	Uso conceptos básicos de probabilidad (espacio muestral, evento, independencia, etc.)	1
5	Pensamiento variacional y sistemas algebraicos y analíticos	2
5.1	Identifico relaciones entre propiedad de las graficas y propiedades de las ecuaciones algebraicas	1
5.2	Construyo expresiones algebraicas equivalentes a una expresión algebraica dada	2
5.3	Uso procesos inductivos y lenguaje algebraico para formular y poner a prueba conjeturas	3
5.4	Modelo situaciones de variación con funciones polinómicas	3
5.5	Identifico diferentes métodos para solucionar sistemas de ecuaciones lineales	1
5.6	Analizo los procesos infinitos que subyacen en las notaciones decimales	2
5.7	Identifico y utilizo diferentes maneras de definir y medir la pendiente de una curva que representa en el plano cartesiano situaciones de variación	2
5.8	Identifico la relación entre los cambios en los parámetros de la representación algebraica de una familia de funciones y los cambios en las graficas que las representan	1
5.9	Analizo en representaciones graficas cartesianas los comportamientos de cambio de funciones especificas pertenecientes a familias de funciones polinómicas, racionales, exponenciales y logarítmicas	2

NDC – Nivel de demanda cognitiva

Anexo 2

Formato para codificar las preguntas

Evaluator: _____

Fecha: _____

Número de pregunta	NDC de la pregunta	Logro Primario	Logro Secundario	Logro Secundario	Comentarios
24					
26					
27					
28					
29					
30					
40					
41					
42					
43					
44					
50					

NDC – Nivel de demanda cognitiva

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Lopez, Alexis A. (alopez@ets.org). Doctor en Educación de la Universidad de Illinois en Urbana-Champaign. Actualmente es investigador en Educational Testing Service – ETS, en Princeton, New Jersey. Anteriormente trabajó como profesor asociado del Centro de Investigación y Formación en Educación –CIFE– de la Universidad de los Andes, en donde también dirigió el Centro de Evaluación. Su dirección postal es Educational Testing Service. 600 Rosedale Road. MS-R04. Princeton, NJ 08901 (USA). [Buscar otros artículos de esta autora en Google Académico / Find other articles by this author in Scholar Google](#)



ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Lopez, Alexis A. (2013). Alineación entre las evaluaciones externas y los estándares académicos: El Caso de la Prueba Saber de Matemáticas en Colombia. <i>RELIEVE</i> , v. 19 (2), art. 3. DOI: 10.7203/relieve.19.2.3024
Title / Título	Alineación entre las evaluaciones externas y los estándares académicos: El Caso de la Prueba Saber de Matemáticas en Colombia. [<i>Alignment between standardized assessments and academic standards: The case of the Saber Mathematics Test in Colombia</i>].
Authors / Autores	Lopez, Alexis A.
Review / Revista	RELIEVE (Revista ELectrónica de Investigación y EValuación Educativa), v. 19, n. 2
ISSN	1134-4032
Publication date / Fecha de publicación	2013 (Reception Date : 2012 November 17 ; Approval Date : 2013 September 17. Publication Date : 2013 September 18)
Abstract / Resumen	<p><i>In this study the concept of alignment is analyzed by presenting the results of a study that examines the degree of alignment between a standardized mathematics assessment and a set of academic standards. The alignment was examined using the Webb model and the results suggest that the degree of alignment between standardized assessments and academic content standards is not the most appropriate and therefore the results should be interpreted with great caution and stakeholders must be very careful with the type of decisions that are made based on the results of these types of assessments.</i></p> <p>En este estudio se analiza el concepto de alineación presentando los resultados de un estudio que examina el grado de alineación entre una evaluación estandarizada de matemáticas y unos estándares académicos. La alineación se examinó usando el modelo de Webb y los resultados sugieren que el grado de alineación entre las evaluaciones estandarizadas y los estándares académicos no es el más adecuado y por lo tanto los resultados se deben interpretar con mucha cautela y se debe poner mucho cuidado con el tipo de decisiones que se toman con base en los resultados.</p>
Keywords / Descriptores	<i>Alignment, Standardized Tests, Academic Standards, Assessment of learning, Validity.</i> Alineación, Pruebas estandarizadas, Estándares Académicos, Evaluación del aprendizaje, Validez.
Institution / Institución	Educational Testing Service (USA).
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español & English version (Title, abstract and keywords in English & Spanish)

RELIEVE

Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).