

# Alineación entre una prueba de aprovechamiento y los estándares curriculares: del discurso a los métodos

## *Alignment between an achievement test and curriculum standards: From discourse to methods*

María del Rosario Medina-Díaz<sup>1</sup>

### Resumen

La alineación entre un instrumento para medir el desempeño estudiantil, como una prueba de aprovechamiento, y los estándares curriculares es uno de los requisitos impuestos por la ley *No Child Left Behind* (Public Law 107-110) y su reemplazo, la ley *Every Student Succeeds* (Public Law 114-95) a los sistemas de evaluación estudiantil de los estados y territorios de los Estados Unidos de América, incluyendo Puerto Rico. Se refiere al grado de acuerdo o armonía entre los ítems de una prueba de aprovechamiento y los estándares curriculares que intenta representar. Este artículo expone brevemente algunos de los métodos tradicionales (*i.e.*, tabla de especificaciones y revisión de conocedoras del contenido) y los particulares, desarrollados por investigadores, para determinar la alineación. Destaca los métodos propuestos por Norman Webb y Andrew Porter y colaboradores. La autora sostiene que el discurso de la alineación de los políticos y funcionarios de las agencias educativas es insuficiente para atender lo que la ley reclama. Deben describir claramente los métodos aplicados y sus resultados. Tampoco la evidencia de la alineación es bastante para reclamar la validez de las puntuaciones de una prueba de aprovechamiento y sus inferencias acerca del aprendizaje estudiantil en las escuelas.

**Palabras clave:** Alineación. Estándares curriculares. Pruebas de aprovechamiento.

### Abstract

*The alignment between a measurement instrument, such as an achievement test, and curriculum standards is one of the requirements for complying with the past No Child Left Behind Act (Public Law 107-110) and the recent Every Student Succeeds (Public Law 114-95) in the states and territories of the United States of America, including Puerto Rico. Alignment refers to the degree to which curriculum standards are in agreement with items of an achievement test. This paper briefly presents some of the traditional (*i.e.*, test blueprint and expert content review) and particular alignment methods (*i.e.*, developed by researchers) that can be applied. It highlights the methods proposed by Norman Webb and Andrew Porter and colleagues. The author argues that the political and educational agency executives' speeches on alignment are*

---

<sup>1</sup> Universidad de Puerto Rico, Facultad de Educación, Departamento de Estudios Graduados. Recinto de Río Piedras, Box 23187, San Juan, Puerto Rico. Correspondencia para/Correspondence to: <medinamaria91@gmail.com>.

*insufficient to sustain the law's requirements. Clearly described methods and results must be demonstrated. Also, the alignment findings are not enough evidence to support the validity of the interpretations and inferences of achievement test scores regarding the students' learning in schools.*

**Keywords:** *Alignment. Curriculum standards. Achievement test.*

### **Alineación entre una prueba de aprovechamiento y los estándares curriculares: Del discurso a los métodos**

La ley conocida como *No Child Left Behind* (*Public Law 107-110*) aprobada en el año 2001 por el Congreso de los Estados Unidos de América, dictó las pautas de una reforma educativa en los sistemas escolares públicos de sus estados y territorios, incluyendo a Puerto Rico<sup>2</sup> por más de una década. Esta ley proclamó un sistema de rendición de cuentas basado en los resultados del desempeño académico del estudiantado, los estándares curriculares y de ejecución<sup>3</sup> (*Public Law 107-110, Title I, Part A, Sec. 1111(b)(1)[A-B]*). Además, dispuso que las agencias educativas desarrollaran y administraran instrumentos de medición (*e.g.*, pruebas de aprovechamiento estandarizadas) a los estudiantes de tercero a octavo y undécimo grado, que estén "alineados" con los estándares de contenido en Matemáticas, Artes del Lenguaje<sup>4</sup> y Ciencia (*Public Law 107-110, Title I, Part A, Sec. 1111(b)(3)(A)[viii]*). La nueva ley *Every Student Succeeds* (*Public Law 114-95*, firmada por el Presidente Barak Obama el 10 de diciembre de 2015) conserva estas exigencias (*Title I, Part B, Sec. 1201(a)(2)(B), [E]*).

El Departamento de Educación de Puerto Rico (DEPR) ha desarrollado estándares curriculares<sup>5</sup> para

estas asignaturas y convertido las pruebas de aprovechamiento estandarizadas (*e.g.*, las Pruebas Puertorriqueñas de Aprovechamiento Académico) en el medio principal para medir la adquisición del conocimiento, las destrezas y las actitudes estipuladas (Medina, 2007). Sus resultados se usan para clasificar el estudiantado en distintos niveles de ejecución y de las escuelas públicas y proveer incentivos o sanciones al personal escolar, a partir de los porcentajes de estudiantes que obtienen el nivel de ejecución fijado en los grados y las asignaturas escolares. Esto ha afectado el énfasis del contenido curricular, el tiempo dedicado a la enseñanza y la honestidad académica, así como la seguridad de empleo y evaluación del magisterio (Hamilton *et al.*, 2007; McMurrer, 2008; Dee *et al.* 2013; Ortíz Franco, 2013; Muñoz Ríos, 2014; Polikoff & Porter, 2014; Jennings & Marc Bearak, 2014; Ballou & Springer, 2015).

Por otra parte, el DEPR ha conservado la exclusividad del uso y la divulgación de los informes de los estudios de alineación de las pruebas de aprovechamiento administradas. Sus funcionarios expresan públicamente que "las pruebas están alineadas" con los estándares. Sin embargo, no señalan cómo se consiguió tal alineación o cuál método se utilizó (López, 2011; Diario de Puerto Rico, 2014). Como sostengo, los resultados de los métodos de alineación

<sup>2</sup> El 25 de julio de 1898 las fuerzas armadas de los Estados Unidos de América invadieron a Puerto Rico. Luego se firmó el Tratado de París entre España y los Estados Unidos de América, que puso fin a la Guerra Hispanoamericana. Desde esa fecha, Puerto Rico es considerado un territorio o una colonia de los Estados Unidos.

<sup>3</sup> Se refieren a los *performance standards* que indican distintos niveles de ejecución (o *proficiency levels*) en los grados escolares, a partir de las puntuaciones de corte establecidas en los instrumentos de medición para cada nivel. En Puerto Rico, los niveles son pre-básico, básico, proficiente y avanzado. Se fijaron con las puntuaciones de las Pruebas Puertorriqueñas de Aprovechamiento Académico, hasta el 2015 y a partir del 2016 con las de Medición y Evaluación para la Transformación Académica de Puerto Rico.

<sup>4</sup> En Puerto Rico, como en los estados y territorios de los Estados Unidos, las pruebas estandarizadas de aprovechamiento incluyen las asignaturas de Matemáticas, Inglés y Ciencia (*Public Law 107-110, Title I, Part A, Sec. 1111(b)(3) Assessments*), pero es el único que incorpora el Español. El idioma de estas pruebas es español, excepto en la prueba de Inglés.

<sup>5</sup> En el año 2013, el Secretario del Departamento de Educación de los Estados Unidos aprobó el *Plan de Flexibilidad* que el Departamento de Educación de Puerto Rico sometió para compensar las dificultades en cumplir con la ley *No Child Left Behind* (Puerto Rico Departamento de Educación, 2013); Puerto Rico, 2015). Al año siguiente, publicó los nuevos estándares curriculares en las asignaturas de Matemáticas, Español, Inglés, Ciencia y Estudios Sociales (disponible en 13 de octubre de 2014).

son diversos y no se reducen al simple discurso de "estar o no alineado"; sino que tiene distintas gradaciones e interpretaciones. Por supuesto, la alineación podría mejorar o empeorar, si hay cambios en los ítems (preguntas o tareas en las pruebas) o en los estándares curriculares. Además, estos métodos no generan evidencia del vínculo entre los distintos estándares curriculares y mucho menos, de las diferencias en los contextos y las condiciones de trabajo en las escuelas públicas.

Por último, conviene señalar que los sistemas nacionales de evaluación en países como Chile, Costa Rica, Argentina, México y Brasil, Inglaterra y Francia también descansan en la administración de pruebas de aprovechamiento estandarizadas para describir el rendimiento escolar (Berlak, 2000; Gvirtz & Larripa, 2004; Wolff, 2006; Stobart, 2010; Calderón & de Oliveira, 2013; Martínez Rizo, 2013; Sánchez Cerón & Corte Cruz, 2013). No obstante, levantan interrogantes acerca de la validez de las puntuaciones y la información provista para orientar el mejoramiento de la enseñanza y el aprendizaje del estudiantado. Muy pocos países, estados o sistemas educativos conducen y publican investigaciones para confirmar y evaluar la calidad técnica y las consecuencias de los instrumentos que desarrollan o que encomiendan a compañías privadas (Medina, 2007; Diaz-Bilello *et al.*, 2014). Además, hay ausencia de evaluaciones sistemáticas del impacto de la mayor parte de las políticas educativas.

### Alineación

El concepto de alineación no es nuevo en el campo de la evaluación y medición educativa. Desde finales de la década del cuarenta, varios autores destacaron la necesidad de la alineación del currículo con otros componentes educativos (Tyler, 1949; Glaser, 1963; Resnick *et al.*, 1973; Crowell & Tissot, 1986; Leitzel & Vogler, 1994). Acuñaron términos como "alineación curricular" (Crowell & Tissot, 1986) y "validez curricular" (Mehrens & Phillips, 1987), para referirse a la correspondencia entre los objetivos curriculares, el currículo y los instrumentos de medición. También la "validez instruccional" y la "sensitividad instruccional"

cobraron importancia para establecer el vínculo entre la ejecución en un instrumento o un ítem y las posibles diferencias en la instrucción recibida (Berk, 1984; Nitko, 1989; Niemi *et al.*, 2007; Popham, 2007; Polikoff, 2010).

Aunque el proceso de alineación podría enfocarse en la armonía entre los diversos componentes de un sistema educativo, se ha concentrado en tres: los estándares de contenido, los instrumentos aplicados para medir el aprovechamiento estudiantil y la instrucción. La alineación entre los primeros dos, que este artículo aborda, es de tipo horizontal. Establece el modo en que los ítems de una prueba de aprovechamiento reflejan los aprendizajes que indican uno o más de los estándares curriculares en un mismo grado o año escolar. Requiere una descripción clara de éstos y la aplicación cuidadosa de un método para establecer la relación.

Los estándares curriculares indican los conocimientos, las destrezas y las actitudes que el estudiantado debe aprender en las distintas asignaturas escolares (*e.g.*, Matemáticas, Inglés, Ciencias, Español). Existe gran variabilidad en la organización, la cantidad y el nivel de especificidad de estos estándares en los sistemas escolares. Por ejemplo, los *Estándares de contenido y expectativas de grado* más recientes del Programa de Matemáticas de Puerto Rico incorporan ocho procesos y competencias fundamentales que el estudiantado debe desarrollar mediante los cursos en esta asignatura, con el fin de prepararlo para los estudios post-secundarios y el mundo profesional (Departamento de Educación de Puerto Rico, 2014, p.2). Además, integra cinco estándares de contenido (Numeración y Operación, Álgebra, Geometría, Medición y Análisis de Datos y Probabilidad), los cuales se descomponen en expectativas e indicadores.

### Métodos de alineación

Existen distintos métodos para establecer el grado de acuerdo o alineación entre los estándares y una prueba de aprovechamiento. Los clasifico en dos grupos: tradicionales o generales y particulares.

### Métodos tradicionales

Los métodos tradicionales se emplean en el diseño de la prueba o luego de construida para asegurar que los ítems representan los estándares, sus expectativas o indicadores. Entre éstos se encuentran la planilla de especificaciones, la revisión de los ítems por personas conocedoras del contenido (conocedoras, de ahora en adelante) y el análisis de documentos. Una planilla o tabla de especificaciones sirve de guía en el diseño y la elaboración de una prueba de aprovechamiento objetiva y de ejecución (Nitko, 1989; Verdejo-Carrión & Medina-Díaz, 2009). La más simple incluiría una columna con los estándares de contenido del grado escolar en cuestión, y otra con la cantidad de ítems que se van a construir o a seleccionar, si hay un banco de ítems disponible. Los estándares se seleccionan mediante la consulta a personas conocedoras del contenido. De este modo, la tabla de especificaciones muestra si existe una alienación mínima (*i.e.*, al menos un ítem representa cada estándar y la mayoría de los ítems emparejan con un estándar, como mínimo).

Además, varios autores recomiendan que se utilice una matriz de doble entrada para precisar la correspondencia entre los ítems, los estándares de contenido y los procesos cognitivos que enfatizan (*e.g.*, Millman, 1989; Anderson, 2002; Lawson & Surtamm, 2006; Osta, 2007; Verdejo-Carrión & Medina-Díaz, 2009). A menudo se recurre a la taxonomía cognoscitiva de Bloom *et al.* (1956), revisada por Anderson y Krathwohl (2001), para establecer las categorías o los niveles de complejidad cognitiva de los estándares y los ítems.

La revisión de los ítems por parte de las conocedoras del contenido es otro modo de percibir la alineación con los estándares curriculares. A un grupo de ellas (*e.g.*, maestras, supervisoras, administradoras escolares y otro personal que labora en los sistemas educativos) se les solicita que indiquen si cada uno de los ítems representan o no alguno de los estándares. Por último, el análisis de documentos también se ha empleado para comprobar la alineación (Webb, 1997). Se presta atención a la concordancia

del contenido de distintos documentos, tales como los estándares curriculares, los libros de texto y los instrumentos de medición.

### Métodos particulares

Los métodos particulares son aquellos que han concretado varios investigadores para alinear los instrumentos de medición el aprovechamiento académico y los estándares curriculares (Porter *et al.*, 1993; Smithson & Porter, 1994; 2004; Gamoran *et al.*, 1997; Webb, 1997, 1999; La Marca *et al.*, 2000; Blank *et al.*, 2001; Porter, 1991, 2002; Porter & Smithson, 2001; Rothman *et al.*, 2002; Resnick *et al.*, 2004; Kulm *et al.*, 2005). Reseño brevemente dos de los más divulgados dentro y fuera de los Estados Unidos, y con referencia a los estándares de Matemáticas. Recomiendo la consulta de las fuentes citadas para abundar en los detalles. Las publicaciones de La Marca *et al.* (2000), Bhola *et al.* (2003), Rothman (2003), Kulm *et al.* (2005), Roach *et al.* (2008) y Martone & Sireci (2009) presentan comparaciones entre los métodos.

### Método de alineación de Webb

El método que Webb (1997; 1999; 2002) propone consiste de dos fases principales: (1) un análisis preliminar del contenido de los estándares curriculares y de los ítems de la prueba y (2) la aplicación de cuatro criterios de alineación - *i.e.*, Concurrencia en las categorías de contenido (*Categorical concurrence*); Consistencia en la profundidad del conocimiento (*Depth of Knowledge [DOK]*); Amplitud del conocimiento (*Range of knowledge*) y Balance de la representación (*Balance of representation*). En la primera fase, cuatro o más conocedoras del contenido asignan a los estándares y sus expectativas uno de los cuatro niveles de profundidad del conocimiento: (1) recordar y reproducir, (2) destreza/concepto, (3) pensamiento estratégico y (4) pensamiento extendido (Webb; 1997; 1999; 2002). Webb (2007) provee una descripción de los procesos cognitivos vinculados con estos niveles. Luego emparejan cada ítem con al menos una expectativa de los estándares curriculares; idealmente,

se espera que cada ítem represente una expectativa. También asignan a cada ítem uno de los niveles DOK. Por último, en una sesión de discusión, comparan los ítems que clasificaron en los distintos DOK, deliberan y logran un consenso.

Una vez concluida esta fase, se aplican los cuatro criterios para determinar la alineación entre los estándares curriculares y la prueba de aprovechamiento en cuestión. Las codificaciones asignadas por cada conocedora del contenido se colocan en una hoja de datos para calcular las estadísticas que reflejan los distintos grados de alienación. El criterio de la concurrencia en las categorías de contenido indica si los ítems de la prueba incorporan las mismas áreas temáticas que los estándares curriculares (e.g., Numeración y Operación, Álgebra). Se identifica el estándar correspondiente a cada ítem y se obtiene la cantidad de ítems asociados. Mientras más ítems se relacionen con un estándar, mayor precisión existe en la medición, y así en la validez en las inferencias derivadas. Webb (2002; 2007) recomienda, como mínimo, seis ítems por cada estándar para cumplir con este criterio. Con cinco ítems, el grado de alineación es débil y un número menor sería inadecuado.

El criterio de la consistencia de la profundidad del conocimiento contrapone la complejidad cognitiva expresada en los estándares curriculares, que requieren los ítems de la prueba. El juicio de las conocedoras acerca del proceso cognitivo implicado se basa en el análisis del contenido de cada uno de los ítems y lo que requiere del estudiantado para contestarlos. Un grado de alineación apropiado requiere que 50% o más de los ítems asignados a las expectativas del estándar tengan el mismo nivel de profundidad del conocimiento.

El criterio de amplitud de conocimiento atiende a la cobertura del contenido de los estándares curriculares; es decir, si el contenido que los ítems abarcan cubre el esperado por los estándares. Se calcula el porcentaje de las expectativas asociadas con cada estándar que están representadas en los ítems. Este criterio se cumple cuando al menos 50%

de las expectativas están ejemplificadas en un ítem o más de la prueba (Webb, 2002). Por lo tanto, supone que el aprovechamiento del estudiantado en los estándares curriculares puede medirse con la mitad de la prueba y que todos los estándares tienen la misma importancia. Aunque se podría establecer mayor rigor en éste y los otros criterios de alineación, la realidad es que es muy difícil de lograrlo cuando hay una gran cantidad de estándares y expectativas y un tiempo limitado para construir la prueba (Webb, 2007).

El cuarto criterio del balance de la representación considera la distribución de los ítems de la prueba entre los estándares curriculares. Presume que los estándares tienen al menos un nivel de especificidad (i.e., expectativa). Se calcula el índice de la distribución considerando la diferencia entre la fracción de las expectativas representadas en un ítem y la de ítems asignadas a cada estándar (Webb, 2007; Martone & Sireci, 2009). Si todos los ítems asignados a un estándar se distribuyen equitativamente entre sus expectativas, el valor del índice de la distribución es uno. Un valor cercano a uno sugiere que la mayoría de los estándares están representados por el mismo número de ítems y, así, la prueba "está balanceada". Mientras mayor sea el número de ítems asociados con un estándar en comparación con los que representan otros, menor será este valor y revela mayor desbalance. Esto apunta a que hay estándares que reciben mayor atención en una prueba. Si el valor es cercano o igual que cero, los ítems representan pocos estándares o su distribución se concentra en uno o dos. Según Webb (2007), los índices de 0,70 o más indican una distribución aceptable de los ítems entre los estándares.

Por último, se agregan los datos que corresponden a cada estándar curricular para crear los perfiles de la alineación. Webb (2002; 2007) señala los valores mínimos, mencionados antes, para ayudar en la interpretación de los resultados. Sin embargo, hay asuntos vinculados con el propósito del instrumento, la importancia de las expectativas de los estándares, el número y el tipo de ítems, la diversidad del contenido en los niveles escolares y las condiciones de administración que deben ponderarse a la hora de tomar decisiones acerca del

desempeño del estudiantado en una prueba de aprovechamiento.

El método de Webb se ha aplicado a distintos instrumentos de medición, en y fuera de los Estados Unidos, para determinar su alineación con ciertos estándares curriculares (Webb, 1999; 2002; 2007; Brown & Conley, 2007; López, 2013; Newton & Kasten, 2013). La página electrónica <<http://wat.wceruw.org/index.aspx>> (denominada *Webb Alignment Tool*) contiene un módulo tutorial y permite la entrada de la codificación. Además, este método se ha modificado y empleado con instrumentos administrados a estudiantes del programa de Educación Especial (Roach & Elliot, 2004; Roach *et al.*, 2005; Elliot & Roach, 2007) y de Educación Pre-escolar (Roach *et al.*, 2010). En Puerto Rico, Webb llevó a cabo un estudio de alineación con las Pruebas Puertorriqueñas de Aprovechamiento Académico de Matemáticas y encontró una alineación pobre con los estándares curriculares (Medina, 2007). La mayoría de los ítems no cumplían con los cuatro criterios de alineación ni presentaban niveles altos de profundidad del conocimiento.

López (2013) también lo aplicó para evaluar la alienación de los ítems de la Prueba Saber de Matemáticas del 2006 y los Estándares Básicos de Competencias de Matemáticas de noveno grado en Colombia. Encontró un bajo grado de alineación entre los ítems de la prueba y los estándares, aun cuando los importantes estaban representados. Brown y Conley (2007) lo utilizaron para determinar la congruencia entre el contenido de las pruebas de aprovechamiento de Artes del Lenguaje en Inglés y Matemáticas que se administraron a los estudiantes de escuela superior en 20 estados y los estándares para el éxito en los estudios universitarios. Concluyeron que la mayoría de las pruebas de Matemáticas no estaban adecuadamente alienadas a estos estándares y que había áreas y niveles cognitivos excluidos.

### **Método *Survey of Enacted Curriculum***

El método *Survey of Enacted Curriculum* es producto de las investigaciones de Andrew Porter y

colaboradores (Porter, 1991; 2002; Porter *et al.*, 1993; Smithson & Porter, 1994; 2004; Gamoran *et al.*, 1997; Blank *et al.*, 2001; Porter & Smithson, 2001). Proponen dos instrumentos para analizar la alineación de una prueba de aprovechamiento con los estándares curriculares: (1) una matriz bidimensional para representar el contenido (*Survey of Instructional Content*) y (2) un cuestionario de prácticas instruccionales (*Survey of Instructional Practices*). La aplicación del método comienza la delimitación de los temas de contenido curricular y las demandas cognitivas. En Matemáticas, por ejemplo, hay identificados, de antemano, 16 áreas y 184 temas para los niveles escolares de Kinder a duodécimo. Cada tema tiene asignado un código numérico (*e.g.*, 101 corresponde al tema de Valor de lugar posicional en el área de Sentido Numérico/ Propiedades/ Relaciones). Las demandas cognitivas, en Matemáticas, se clasifican en cinco categorías: Memorizar hechos, definiciones y fórmulas; Ejecutar procedimientos; Demostrar entendimiento de ideas matemáticas; Realizar conjeturas/Generalizar/Probar; y Solucionar problemas no rutinarios/ Establecer conexiones. Ambos se colocan en una matriz de contenido instruccional. La dimensión horizontal podría incluir las áreas de contenido (*coarse grain*) o los temas específicos de cada área (*fine grain*) y la vertical, las demandas cognitivas.

Cuatro o más conocedoras del contenido analizan individualmente los estándares curriculares y los ítems de la prueba. Las celdas de matriz del contenido instruccional sirven de marco para codificarlos de la manera más específica posible. Se espera que ubiquen cada expectativa de los estándares en una o más de las celdas de la matriz, dependiendo de su complejidad. Asimismo ocurre con cada ítem de la prueba. La celda en que se colocan los ítems va acumulando una puntuación relativa. Si a un ítem se le asigna un punto y se ubica en dos celdas, se distribuye esta puntuación y correspondería 0,50 a cada celda. Lo mismo aplica a un ítem que tiene una puntuación mayor (Polikoff *et al.*, 2011). Luego, las conocedoras se reúnen para discutir las dudas en la codificación de algunos de los

ítems. Como resultado, podrían cambiar su codificación inicial.

Así se obtiene un promedio aritmético de las puntuaciones de cada celda y se produce una matriz de fracciones, la cual se utiliza para calcular el índice de alineación. Las fracciones son un estimado del énfasis relativo del contenido de cada celda en los estándares y en la prueba. Se calcula un índice de alineación que registra el grado en que ambos cubren el mismo contenido (Porter, 2002 para la fórmula). Su valor fluctúa entre 0 y 1, donde uno indica una alineación perfecta (*i.e.*, que ambos documentos tienen 100% del contenido en común). Según Smithson y Porter (2004) se interpreta como un coeficiente de correlación, pero es difícil de determinar el valor óptimo. Fulmer (2011) propone un procedimiento estadístico mediante simulaciones para determinar los valores críticos y comparar los obtenidos.

Además, los resultados se exhiben en una gráfica de barras o un mapa topográfico. Ilustra la mayor o menor frecuencia (*i.e.*, énfasis) de las áreas (o los temas) de contenido y las demandas cognitivas. En las publicaciones de Smithson y Porter (2004); Porter *et al.* (2011); Porter *et al.* (2013) y en el tutorial del *Survey of Enacted Curriculum* se encuentran varios ejemplos de mapas topográficos <<http://seconline.wceruw.org>>. El contorno y la intensidad de los colores muestran las diferencias en el contenido y las demandas cognitivas de cada documento analizado. Aquí no se comparan directamente los ítems de una prueba de aprovechamiento con los estándares, sino las matrices que los representan.

El *Survey of Enacted Curriculum*, a diferencia del método de Webb, combina el análisis del contenido de los estándares y la prueba de aprovechamiento con información de lo que actualmente se enseña en los salones de clase (*i.e.*, énfasis instruccional), mediante la administración de un cuestionario a un grupo de maestras y maestros. El cuestionario indaga acerca de los temas de contenido que cubrieron o enseñaron, en un periodo de tiempo determinado. Se les solicita que indiquen el tiempo dedicado a la enseñanza de cada uno de los temas (*i.e.*, nivel de cobertura) y el énfasis relativo que otorgaron a las categorías de demandas cognitivas en cada uno de

los temas cubiertos. Incluye un conjunto de preguntas acerca de su práctica instruccional. Las maestras y los maestros pueden contestarlo en cualquier momento del año escolar, aunque Porter (2002) recomienda que sea al final. Así, se estima el tiempo y énfasis dedicado a cada tema y a las categorías de demanda cognitiva. Estos datos también se colocan en una matriz de contenido instruccional y se calculan las fracciones correspondientes.

Finalmente, los resultados de la codificación de los tres documentos examinados (*i.e.*, prueba de aprovechamiento, estándares curriculares y cuestionario de prácticas instruccionales) se colocan en una matriz de contenido instruccional y se comparan para hallar el índice de alineación. Los valores iguales a uno (*i.e.*, alineación perfecta) son casi imposibles de obtener. En la práctica, estos valores no sobrepasan a 0,50 (Polikoff *et al.*, 2011; Porter *et al.*, 2013). Porter y colaboradores (Porter & Smithson, 2001; Porter *et al.*, 2007; 2008; 2011; 2013; Smithson & Porter, 2004; Blank & Smithson, 2009; Polikoff *et al.*, 2011) han aplicado el *Survey of Enacted Curriculum*, contrastando prácticas instruccionales en los salones de clases, contenido de los estándares curriculares e instrumentos administrados en estados y distritos escolares de los Estados Unidos. Para ilustrar, Polikoff *et al.* (2011) examinaron los datos del *Survey of Enacted Curriculum* sobre la alineación entre los estándares de contenido en Inglés (de 11 estados), Matemáticas (de 14) y Ciencias (de nueve) y los respectivos instrumentos de medición. Los índices de alineación fluctuaron entre 0,20 y 0,30. Encontraron que el promedio aritmético de este índice en Matemáticas (0,27) y Ciencia (0,26) era un poco más alto que el de Inglés (0,19). Recurrieron a los mapas topográficos para interpretar los resultados. En Matemáticas, señalaron que, a pesar de la baja magnitud del índice (0,27), los mapas revelaban que había un traslape sustancial en algunos de los temas de contenido y la demanda cognitiva de ejecutar procedimientos (Polikoff *et al.*, 2011, las figuras 1 y 2).

## Conclusión

La alineación entre una prueba de aprovechamiento y los estándares curriculares

procura determinar el grado de acuerdo entre ambos. No se trata de una cualidad "que tiene" una prueba o los estándares curriculares, sino de la armonía que existe entre éstos. Para lograrlo, las agencias educativas de los estados y países necesitan establecer claramente los estándares curriculares, los propósitos de los estudios de alineación y el uso de los resultados. Asimismo, la credibilidad y la efectividad de la alineación dependen de la colaboración de las personas conocedoras del contenido para revisar estos documentos con suficiente tiempo y rigor. Por lo tanto, es necesario adiestrarlas y estipular la consistencia de sus clasificaciones, calculando las estadísticas apropiadas.

Subrayo que es imposible una relación perfecta entre una prueba de aprovechamiento y los estándares curriculares. Es poco probable que ésta u otro instrumento de medición pueda capturar la amplitud y profundidad del contenido, así como la complejidad de los procesos cognitivos implicados, y lograr un grado de alineación impecable con los estándares curriculares y sus expectativas. Entre las razones para esto están los criterios aplicados a los métodos de alineación, la elección, la calidad del adiestramiento de las conocedoras del contenido al aplicarlos, la ambigüedad y cobertura de los estándares y de la colección de los ítems de la prueba de aprovechamiento. Los resultados de los estudios mencionados revelan debilidades en la alineación deseada.

Aun cuando se espera que los ítems cubran todos los estándares curriculares, en la práctica unos reciben mayor importancia que otros, según el nivel escolar. En el primario, por ejemplo, los estándares de Numeración y Operación tienen mayor relevancia. Lo crítico es que la prueba refleje las variaciones en la importancia y la complejidad de los estándares, por grados. Por supuesto, esto está atado a los controles en el tiempo de administrarla que limita el número de ítems que se pueden incluir.

Los métodos de alineación reseñados coinciden en la búsqueda de la concordancia entre el contenido de los estándares curriculares y los ítems de una prueba. Difieren en el examen de las demandas

cognitivas y en los criterios y procedimientos empleados. Tanto el método de Webb como el de Porter y colaboradores producen diferentes resultados de la alineación, según los criterios fijados y las interpretaciones provistas (Newton & Kasten, 2013 para una comparación).

El método de Webb (1997; 1999) provee varios índices cuantitativos de la afinidad entre los ítems y los estándares curriculares conforme a los cuatro criterios propuestos. Webb (1999) presenta ciertas guías para interpretarlos, al momento de este escrito no se ha publicado un índice de alineación general. Por el contrario, el *Survey of Enacted Curriculum* produce uno y una representación gráfica asociada (*i.e.*, mapa topográfico). Sin embargo, no establece indicaciones para apreciar si es óptimo. Se interpreta a la luz de los documentos analizados y los mapas topográficos.

Por otra parte, ninguno de estos métodos considera la alineación vertical para comparar el aprovechamiento del estudiante entre los grados escolares y la posible aportación del estudiantado al proceso de alineación, al menos para auscultar su punto de vista acerca de la enseñanza basada en los estándares curriculares (Roach *et al.*, 2008) o de la efectividad de los ítems en representar los aprendizajes esperados (Kulm *et al.*, 2005). El estudio de Kulm *et al.*, 2005, aplicando un procedimiento de alineación desarrollado en el Proyecto 2016 de la Asociación Americana para el Avance de las Ciencias, constituye un paso de avance en esta dirección. Entrevistaron a 13 estudiantes de cuarto grado y siete de sexto para indagar acerca del proceso de contestar y de la interpretación y el contexto de dos tareas de ejecución de Matemáticas. Hallaron que sus pensamientos no correspondían a los que solicitaban los estándares.

En conclusión, el uso de una prueba de aprovechamiento escolar en un estado o país invita a una mayor atención a la construcción y la validez de las interpretaciones y las inferencias de sus puntuaciones para determinar los niveles de desempeño del estudiantado. La alineación con los estándares curriculares es parte de la evidencia acerca



del contenido, pero insuficiente para garantizar y sostener el argumento de la validez del uso de las puntuaciones. Requiere acumular evidencia vinculada a distintas fuentes, tales como el proceso de respuesta, la estructura interna, la relación con otras variables y de las consecuencias (Messick, 1989; Kane, 2002; Medina, 2007; Medina Díaz, 2013; American Educational Research Association *et al.*, 2014). Además, se necesita que la instrucción, los recursos humanos, los materiales educativos, el apoyo familiar e institucional y las condiciones de trabajo en las escuelas, entre otros, fomenten el aprovechamiento estudiantil.

Recomiendo, por lo tanto, examinar con cuidado la documentación acerca de la alineación que proveen las compañías o las personas que construyen estas pruebas. El método y los criterios de alineación, las características de las conocedoras del contenido y la interpretación de los resultados de la alienación demandan más miramiento que el discurso de los políticos y los funcionarios de los sistemas educativos. Una práctica contraria tiene repercusiones en la validez de las puntuaciones y de las inferencias asociadas con el rendimiento escolar. Con una alineación pobre entre una prueba de aprovechamiento y los estándares curriculares, una agencia educativa en un estado o territorio está en una posición muy débil para cumplir con lo que una ley impulsa. Peor aún, conduce a emitir juicios errados acerca del aprovechamiento y la enseñanza y a tomar decisiones que perjudican las escuelas, la clase magisterial y el estudiantado.

## Referencias

- American Educational Research Association; American Psychological Association; National Council on Measurement. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association, 2014.
- Anderson, L.A. Curricular alignment: A re-examination. *Theory into Practice*, v.41, n.4, p.255-260, 2002.
- Anderson, L.A.; Krathwohl, D.R. (Ed.). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.
- Ballou, D.; Springer, M.G. Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, v.44, n.2, p.77-86, 2015.
- Berk, R. (Ed.). *A guide to criterion-reference test construction*. Baltimore, MD: Johns Hopkins University Press, 1984.
- Berlak, H. Cultural politics, the science of assessment and democratic renewal of public education. In: Filer, A. (Ed.). *Assessment: Social practice and social product*. New York: Routledge, 2000. p.189-207.
- Bhola, D.S.; Impara, J.C.; Buckdenhal, C.W. Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, v.22, p.21-29, 2003.
- Blank, R.K.; Smithson, J. *Alignment content analysis of TIMSS and PISA Mathematics and Science assessments using the Surveys of Enacted Curriculum methodology*. Washington, DC: Council of Chief State School Officers, 2009.
- Blank, R.K.; Porter, A.; Smithson, J. *New tools for analyzing teaching, curriculum and standards in Mathematics and Science*. Washington, DC: Council of Chief State School Officers, 2001. (Results from Survey of Enacted Curriculum Project: Final Report).
- Bloom, B.S. *et al. Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay, 1956. (Handbook I: Cognitive Domain).
- Brown, R.; Conley, D. Comparing state high school assessments to standards for success in entry-level university courses. *Journal of Educational Assessment*, v.12, p.137-160, 2007.
- Calderón, A.I.; de Oliveira, R. G. Sistema de Avaliação e rendimento escolar do estado de São Paulo. In: Martins, A.M., *et al.* (Org.). *Políticas e gestão da Educação: desafios em tempos de mudanças*. Campinas: Autores Associados, 2013. p.139-163.
- Crowell, R.; Tissot, P. *Curriculum alignment*. Washington, DC: Office of Educational Research and Improvement, 1986. (ERIC Document Reproduction Service nº 280874).
- Dee, T.S.; Jacob, B.; Schwartz, N.L. The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, v.35, n.2, p.252-279, 2013.
- Departamento de Educación de Puerto Rico. *Estándares de contenido y expectativas de grado*. San Juan: Departamento de Educación de Puerto Rico, 2014. (Programa de Matemáticas).
- Departamento de Educación de Puerto Rico. *Plan de Flexibilidad ESEA*, 2013. Disponible en: <[http://www.de.gobierno.pr/files/Plan\\_Flexibilidad\\_2014\\_ aprobado.pdf](http://www.de.gobierno.pr/files/Plan_Flexibilidad_2014_ aprobado.pdf)>. Acceso en: 8 feb. 2015.
- Diario de Puerto Rico. Prioridad para el DE que resultados de las Pruebas Puertorriqueñas reflejen crecimiento académico individual de los estudiantes, 22 de agosto de

2014. Disponible en: <<http://diariodepuertorico.com/2014/08/prioridad-para-el-de-que-resultados-de-las-pruebas-puertorriquenas-reflejen-crecimiento-academico-individual-de-los-estudiantes/>>. Acceso: 2 jun. 2015.
- Diaz-Bilello, E. et al. Are the Standards for Educational and Psychological Testing relevant to state and local assessment programs? *Educational Measurement: Issues and Practice*, v.33, n.4, p.16-18, 2014.
- Elliot, S.; Roach, A.T. Alignment of students with significant disabilities: Alternative approaches, common technical challenges. *Applied Measurement in Education*, v.20, n.3, p.301-333, 2007.
- Fulmer, G.W. Estimating critical values for strength of alignment among curriculum, assessments, and instruction. *Journal of Educational and Behavioral Sciences*, v.36, p.381-402, 2011.
- Gamoran, A. et al. Upgrading high school mathematics instruction: Improving learning opportunities for low-achievement, low-income youth. *Educational Evaluation and Policy Analysis*, v.19, p.325-338, 1997.
- Glaser, R. Instructional technology and the measurement of learning out comes: Some questions. *American Psychologist*, v.18, p.519-521, 1963.
- Gvirtz, S.; Larripa, S. National evaluation system in Argentina: Problematic present and uncertain future. *Assessment in Education*, v.11, n.3, p.349-364, 2004.
- Hamilton, L.S. et al. *Standard-based accountability under no child left behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: Rand, 2007.
- Jennings, J.L.; Marc Bearac, J. "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, v.43, n.8, p.381-389, 2014.
- Kane. M.T. Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, v.21, n.1, p.31-41, 2002.
- Kulm, G.; Wilson, L.D.; Kitchen, R. Alignment of content effectiveness of Mathematics assessment items. *Educational Assessment*, v.10, n.4, p.333-356, 2005.
- La Marca, P.M., Redfield, P.C.; Winter, A. *State standards and state assessment systems: A guide to alignment*. Washington, DC: Council of Chief State School Officers, 2000.
- Lawson, A.; Surtamm, C. The challenges and possibilities of alignment in large-scale testing with mathematical reform: The case of Ontario. *Assessment in Education*, v.13, n.3, p.305-325, 2006.
- Leitzel, T.C.; Vogler, D.E. *Curriculum alignment: Theory and practice*. 1994. (ERIC Document Reproduction Service n.371812). Disponible en: <<http://files.eric.ed.gov/fulltext/ED371812.pdf>>. Cited: 15 Jul. 2011.
- López, A. A. Alineación entre las evaluaciones externas y los estándares académicos: El caso de la Prueba Saber de Matemáticas en Colombia. *Relieve*, v.19, n.2, 2013, <<http://dd.doi.wg/10.7203/relieve.19.1.3024>>.
- López Alicea, K. Todos tienen que rendir cuentas, dice el Secretario de DE. *El Nuevo Día*, 2011, p.2. Disponible en: <<http://www.elnuevodia.com/aquitenemosquerendir-cuentas-897939.html>>. Acceso: 1 mar. 2011.
- Martínez Rizo, F. El futuro de la evaluación educativa. *Sinéctica*, v. 40, 2013. Disponible en: <[http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=591\\_el\\_futuro\\_de\\_la\\_evaluacion\\_educativa](http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=591_el_futuro_de_la_evaluacion_educativa)>. Acceso: 7 feb. 2015.
- Martone, A.; Sireci, S.G. Evaluating alignment between curriculum, assessment and instruction. *Review of Educational Research*, v.79, p.1332-1361, 2009.
- McMurrer, J. *Instructional time in elementary schools: A closer look at the changes for specific subjects, from the capital to the classroom: Year 5 of the No Child Left behind Act*. Washington, DC: Center on Educational Policy, 2008.
- Mehrens, W.A.; Phillips, S.E. Sensitivity of item difficulties to curricular validity. *Journal of Educational Measurement*, v.24, n.4, p.357-370, 1987.
- Medina Díaz, M. del R. ¿"Validez" o "validación" de un instrumento? *INEVA en Acción*, v.9, n.1, p.1-4, 2013. Disponible en: <<http://ineva.uprrp.edu>>. Acceso: 7 feb. 2015.
- Medina, M.R. Las pruebas de aprovechamiento estandarizadas como instrumento de medición y político. *Pedagogía*, v.40, n.1, p.145-187, 2007.
- Messick, S. Validity. In: Linn, R.L. (Ed.). *Educational measurement* (3rd. Ed.). New York: American Council on Education and MacMillan, 1989. p.13-103.
- Millman, J.; Green, J. The specification and development of tests of achievement and ability. In: Linn, R.L. (Ed.). *Educational measurement* (3<sup>rd</sup>. ed.). Phoenix, AZ: Oryx Press, 1989. p.335-366.
- Muñoz Ríos, M. *Actividades de las maestras de español de cuarto a sexto grado para preparar a los estudiantes para contestar las pruebas estandarizadas y el uso de los resultados*. Tesis (Maestría inédita) - Programa de maestría en Investigación y Evaluación Educativa. Universidad de Puerto Rico, Recinto de Río Piedras, 2014.
- Newton, J.A.; Kasten, S.E. Two models for evaluating alignment of state standards and assessments: Competing or complementary perspective? *Journal for Research in Mathematics Education*, v.44, p.550-581, 2013.
- Niemi, D. et al. Instructional sensitivity of a complex language arts performance assessment. *Educational Assessment*, v.12, p.215-237, 2007.
- Nitko, A.J. Designing tests that are integrated with instruction. In: Linn, R.L. (Ed.). *Educational measurement*. 3<sup>rd</sup>. ed. New York: MacMillan, 1989. p.447-474.

- Ortíz Franco, E. *Prácticas educativas de las(os) maestras(os) de Matemática y la percepción sobre las pruebas estandarizadas de aprovechamiento académico*. Tesis (Maestría inédita) - Programa de maestría en Investigación y Evaluación Educativa, Universidad de Puerto Rico, Recinto de Río Piedras, 2013.
- Osta, I. Developing and piloting a framework for studying the alignment of Mathematics Examinations with the curriculum: The case of Lebanon. *Educational Research and Evaluation*, v.13, n.2, p.171-198, 2007.
- Polikoff, M.S. Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice*, v.29, n.4, p.3-14, 2010.
- Polikoff, M.S.; Porter, A.C. Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, v.36, n.4, p.399-416, 2014.
- Polikoff, M.S.; Porter, A.C.; Smithson, J. How well aligned are the state assessment of student with content standards? *American Educational Research Journal*, v.20, p.1-31, 2011.
- Popham, W.J. Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan*, v.89, n.2, p.146-155, 2007.
- Porter, A.C. Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, v.13, n.1, p.13-29, 1991.
- Porter, A.C. Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, v.31, n.7, p.3-14, 2002.
- Porter, A.C. et al. Alignment as a teacher variable. *Applied Measurement in Education*, v.20, n.1, p.27-51, 2007.
- Porter, A. et al. Common core standards: The new U.S. intended curriculum. *Educational Researcher*, v.40, n.3, p.103-116, 2011.
- Porter, A. et al. Constructed alignment assessments using automated test construction. *Educational Researcher*, v.42, p.415-423, 2013.
- Porter, A. et al. *Reform Up Close: An Analysis of High School Mathematics and Science Classrooms*. Madison, WI: University of Wisconsin-Madison, 1993.
- Porter, A.C. et al. The quality of content analysis of state student achievement tests and content standards. *Educational Measurement: Issues and Practice*, v.27, n.4, p.2-14, 2008.
- Porter, A.C.; Smithson, J.L. *Defining, developing, and using curriculum indicators*. Philadelphia, PA: Consortium for Policy Research in Education, 2001.
- Puerto Rico Department of Education. *Flexibility Request 2015-2018*. Puerto Rico: Departamento de Education. Available from: <[http://www.de.gobierno.pr/files/PRDE\\_Flexibility\\_Request\\_2015\\_2018.pdf](http://www.de.gobierno.pr/files/PRDE_Flexibility_Request_2015_2018.pdf)>. Cited: 31 Aug. 2015.
- Resnick, L.B. et al. Benchmarking and alignment of standards and testing. *Educational Assessment*, v.9, p.1-27, 2004.
- Resnick, L.B.; Wang, M.C.; Kaplan, J. Task analysis in curriculum design: A hierarchically sequenced introductory Mathematics curriculum. *Journal of Behavioral Analysis*, v.6, p.679-710, 1973.
- Roach, A.T.; Elliot, S.N. *Alignment analysis and standard setting procedures for alternate assessments*. Madison: University of Wisconsin, 2004. (WCER Working Paper nº 2004-1).
- Roach, A.T.; Elliot, S.N.; Webb, N.L. Alignment of an alternative assessment with state academic standards: Evidence for the content validity of the Wisconsin alternate assessment. *Journal of Special Education*, v.38, p.218-231, 2005.
- Roach, A.T. et al. Aligning an early childhood assessment to state kindergarten content standards: Application of a nationally recognized alignment framework. *Educational Measurement: Issues and Practice*, v.29, p.25-37, 2010.
- Roach, A.T.; Niebling, B.C.; Kurz, A. Evaluating the alignment among curriculum, instruction, and assessments: Implications and applications for research and practice. *Psychology in the Schools*, v.45, n.2, p.158-176, 2008.
- Rothman, R. *Imperfect matches: The alignment of standards and tests*. Washington, DC: National Research Council, 2003.
- Rothman, R. et al. *Benchmarking and alignment of standards and testing*. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, 2002. (CSE Technical Report nº 566).
- Sánchez Cerón, M.; Corte Cruz, F.M. Las evaluaciones estandarizadas: sus efectos en tres países latinoamericanos. *Revista Latinoamericana de Estudios Educativos*, v.43, n.1, p.97-124, 2013.
- Smithson, J.L.; Porter, A.C. From policy to practice: The evolution of one approach to describing and using curriculum data. In: Wilson, M. (Ed.). *Towards Coherence between classroom assessment and accountability*. Chicago, IL: University of Chicago, 2004. p.105-131.
- Smithson, J.S.; Porter, A.C. *Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum*. Madison, WI: Consortium for Policy Research in Education, 1994.
- Stobart, G. *Tiempo de pruebas: los usos y abusos de la evaluación*. Madrid: Ediciones Morata, 2010.
- Tyler, R.W. *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago, 1949.
- Verdejo-Carrión, A.L.; Medina-Díaz, M. del R. *Evaluación del aprendizaje estudiantil*. (Ed.). San Juan: ExPERTS Consultants, 2009.
- Webb, N.L. *Criteria for alignment of expectations and assessments in Mathematics and Science education*. Madison,

WI: University of Wisconsin-Madison; National Institute of Science Education, 1997. (NISE Research Monograph nº 6).

Webb, N.L. *Alignment of Science and Mathematics standards and assessments in four States*. Madison, WI: University of Wisconsin-Madison; National Institute of Science Education, 1999. (NISE Research Monograph nº 18).

Webb, N.L. An analysis of the alignment between Mathematics standards and assessments for three states. In: *American Educational Research Association Annual Meeting*. Madison, WI: University of Wisconsin-Madison, 2002.

Webb, N.L. Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, v.20, n.1, p.7-25, 2007.

Wolff, L. Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos. In: Arregi, P. (Ed.). *Sobre estándares y evaluaciones en América Latina*. Santiago: Programa de Promoción de Reforma Educativa en América Latina y el Caribe, 2006. p.13-52.

Recibido en 29/7/2015 y aceptado para su publicación en 16/6/2016.