

Para citar este artículo, le recomendamos el siguiente formato:
Gallardo, K., Valenzuela, J. (2014). Evaluación del desempeño: acercando la investigación educativa a los docentes. Revista de evaluación educativa, 3 (2). Consultado el día de mes de año en:
<http://revalue.mx/revista/index.php/revalue/issue/current>

Evaluación del desempeño: acercando la investigación educativa a los docentes

Katherina Edith Gallardo Córdova
Instituto Tecnológico y de Estudios Superiores de Monterrey, México.

Jaime Ricardo Valenzuela González
Instituto Tecnológico y de Estudios Superiores de Monterrey, México.

Resumen

Este artículo presenta los resultados de un proyecto de investigación orientado a diseñar y aplicar un sistema de evaluación del desempeño académico, para promover la mejora de la calidad educativa en instituciones públicas de educación básica. Se siguió una metodología *ex-post-facto*, consistente en aplicar 20 exámenes, para cuatro disciplinas, en 105 escuelas del estado de Tabasco (México). Los resultados muestran información sobre los niveles de desempeño académico de los estudiantes participantes; y un conjunto de lecciones aprendidas acerca de la forma en que profesores en ejercicio realizan procesos de evaluación del aprendizaje.

Palabras clave: Evaluación del aprendizaje, desempeño académico, taxonomía, objetivos de aprendizaje, diseño de instrumentos, educación basada en competencias

Abstract

This article presents the results of a research project aimed at designing and implementing a system for evaluating academic performance , to encourage improvements in the quality of education in public institutions of basic education. An *ex -post- facto* , applied methodology of 20 tests, four disciplines in 105 schools in the state of Tabasco (Mexico) was followed. The results show about levels of academic performance of participating students ; and a set of lessons learned about how practicing teachers made learning assessment processes .

Keywords: Learning evaluation, academic performance, taxonomy, learning objectives, instrument design, competency-based education.

Fecha de recepción: 12 de enero 2015
Fecha de aceptación: 29 de enero 2015

Introducción

La evaluación educativa es un elemento fundamental para mejorar la calidad educativa. A pesar de lo anterior, la historia que tienen los procesos estandarizados de evaluación del aprendizaje en México es relativamente reciente, no más de 25 años. La responsabilidad social y económica del Estado para promover los mecanismos de evaluación a nivel nacional ha hecho que se propongan iniciativas tanto para mejorar la forma de evaluar, como para el proceso de toma de decisiones a partir de los resultados obtenidos (Vidal, 2009).

En México, uno de los primeros esfuerzos de evaluación del aprendizaje de forma censal es el que se refiere a la Evaluación Nacional del Logro Académico en Centros Escolares, conocida también como la Prueba ENLACE. Esta prueba se aplicó por primera vez en el año 2006 a estudiantes de educación básica (primaria y secundaria) de todas las escuelas, públicas y privadas, en el país. La Prueba ENLACE está orientada a evaluar el desempeño de los alumnos en matemáticas, español y otras disciplinas (ciencia, formación cívica y ética e historia) que varían de un año a otro. Los resultados arrojados por esta evaluación se han convertido en un referente importante sobre niveles de logro y rendición de cuentas de la educación. Es a partir de estos resultados que los estados de la República han identificado, de una manera más precisa, sus fortalezas y debilidades, así como las necesidades para mejorar la educación.

Un caso particular es el estado de Tabasco, el cual se encuentra ubicado en el sureste de la República Mexicana. Los resultados desfavorables en ENLACE, obtenidos en los últimos años por esta entidad, han orientado los esfuerzos de la Secretaría de Educación de Tabasco (SETAB) y del mismo gobierno estatal a buscar alternativas para elevar la calidad educativa. Una de estas alternativas consistió en solicitar, a través del Consejo Nacional de Ciencia y Tecnología (CONACYT), fondos para una investigación orientada a desarrollar un sistema de evaluación externo que validara los resultados de la Prueba ENLACE y diera información más contextualizada sobre la naturaleza de la educación en la entidad. Para la asignación de fondos se realizó un concurso, ganando el proyecto la Escuela de Graduados en Educación del Tecnológico de Monterrey.

Objetivo de la investigación

El proyecto de investigación tuvo como objetivo diseñar y aplicar un sistema de evaluación del desempeño académico, distinto a los aplicados de manera oficial, y que sirva para promover la mejora de la calidad educativa en las instituciones públicas de educación básica del estado de Tabasco.

En este contexto, el concepto de *sistema* (definido así por la convocatoria del CONACYT) implicó el establecer un conjunto de entradas, un conjunto de salidas y un proceso entre ellas. Las entradas del sistema consistieron en definir un conjunto de instrumentos orientados a medir desempeño académico, así como

una muestra representativa de estudiantes por ser evaluados. Las salidas del sistema consistieron en los informes acerca del desempeño académico de los estudiantes participantes en el estudio. El proceso correspondió a la aplicación de los instrumentos de evaluación a la muestra seleccionada, así como al análisis de los datos colectados. El sistema desarrollado para este proyecto contrasta fundamentalmente con la prueba ENLACE en el sentido de que, mientras el primero se acotó a una muestra de un estado (Tabasco) de la República Mexicana, ENLACE se aplica de manera censal a todas las escuelas de educación básica de todo el país.

El proyecto fue dividido en tres etapas: (1) el diseño de instrumentos, (2) la aplicación de éstos en escuelas de educación básica y (3) la capacitación a maestros en diferentes aspectos relacionados con el proceso de enseñanza-aprendizaje. Para los propósitos de este artículo se reportan los resultados de la primera y segunda etapa.

Referentes conceptuales

En los últimos años se ha puesto mucho énfasis en el concepto de Educación Basada en Competencias (EBC) en los ambientes educativos. México ha adoptado este modelo educativo a tal punto que las Reformas Integrales de Educación planteadas en los últimos cinco años apuntan a la transformación de un sistema tradicional hacia el que marca el modelo por competencias (<http://basica.sep.gob.mx/reformaintegral/sitio/>). Desde que tuvo lugar la inserción del modelo de EBC en México, los profesores se han enfrentado a la aparente incompatibilidad del uso de objetivos de aprendizaje con este nuevo modelo educativo. La taxonomía de Bloom (Bloom, Engelhart, Furst, Hill y Krathwohl, 1956) ha estado muy presente en el sistema educativo mexicano para redactar los objetivos de aprendizaje de las distintas disciplinas. El uso de esta taxonomía ha sobrevivido a distintas reformas educativas en el país y parece ser que no ha importado que dicha taxonomía tenga más de cinco décadas de existencia. Con la implementación de un modelo de EBC, a los profesores se les solicita que dejen de lado la tradicional taxonomía de Bloom y que piensen en las competencias que sus alumnos deben desarrollar dentro de las disciplinas, como si el propósito de desarrollar dichas competencias no fuera ya un objetivo en sí. El rompimiento con la taxonomía de Bloom es palpable y, al mismo tiempo, al profesor no se le brinda un marco de referencia para organizar y jerarquizar las competencias de acuerdo con el nuevo modelo.

Parte de la confusión que muchos profesores tienen respecto a los modelos de EBC proviene de una falta de claridad semántica respecto al uso de ciertos términos, como lo son aprendizaje, conocimiento y competencia. En el caso de México, esta falta de claridad se debe en buena medida a que mucho del desarrollo de políticas públicas parece poner más énfasis en cuestiones curriculares, dejando de lado principios de psicología educativa. Esto, de hecho, no se hace explícito en documentos oficiales (e.g.,

<http://basica.sep.gob.mx/reformaintegral/sitio>) que carecen de referencias que permitan sustentar la manera en que se está enfocando el modelo basado en competencias, pero sí se puede evidenciar si observamos el discurso de ponentes invitados a congresos, organizados a nivel federal y estatal, para hablar de estos temas (e.g., Ruiz, 2009 y Zabala, 2010). La falta de sustento en principios psicopedagógicos ha hecho que, en México, algunos tiendan a distinguir el conocimiento de las habilidades, reduciendo el primer término a aspectos puramente memorísticos; mientras que el concepto de competencia está primordialmente relacionado con las habilidades – con la capacidad de las personas para hacer algo. Esta visión contrasta con la que aporta la psicología cognitiva, en la que el conocimiento es un constructo que involucra tres elementos: el conocimiento declarativo, vinculado con la memoria; el conocimiento procedimental, relacionado con las habilidades; y el conocimiento condicional, asociado con el contexto en el que cierto conocimiento puede ser aplicable (Woolfolk, 2009). En este enfoque, la competencia de una persona se define como su capacidad para usar el conocimiento declarativo, procedimental y condicional en una situación determinada. El aprendizaje definido como la adquisición de conocimiento, se podría ver como sinónimo del desarrollo de competencias.

La taxonomía de Bloom es una propuesta que surge en los inicios de la psicología cognitiva de los Estados Unidos a mediados de los años 1950s. De ese entonces a la fecha, la psicología cognitiva ha hecho enormes avances. Una de las aportaciones más recientes en este campo es la que hacen Marzano y Kendall (2007) y que dio como resultado el libro *The new taxonomy of educational objectives* (La nueva taxonomía de objetivos educativos). En ella se incorporan distintas aportaciones de la psicología cognitiva que sustentan una nueva clasificación consistente en seis niveles de procesamiento y tres dominios del conocimiento. Por un lado, los niveles de procesamiento se refieren a tres sistemas: (1) el *self*, que se refiere a la motivación para el aprendizaje; (2) el metacognitivo, que explica los procesos de supervisión y determinación de metas de aprendizaje; y (3) el cognitivo, que incluye los procesos de recuperación de información, comprensión, análisis y utilización del conocimiento. Por otro lado, los dominios de conocimiento se refieren a: (1) información, que es el conjunto organizado de datos que se tienen en la memoria; (2) procedimiento mentales, que corresponde al manejo de información que las personas realizan para resolver problemas, tomar decisiones, emitir juicios críticos, etc.; y (3) procedimientos psicomotores, que involucran la coordinación de los procesos mentales con movimientos corporales con diferentes finalidades. Esta taxonomía ha sido empleada extensamente en diseño de ambientes de aprendizaje, procesos de enseñanza-aprendizaje y sistemas de evaluación (Marzano y Kendall, 2008).

Las ideas anteriores establecen tres premisas: (1) el concepto de competencia está íntimamente vinculado con el concepto amplio que la psicología cognitiva contemporánea da al constructo *conocimiento*; (2) el diseño de programas educativos en un modelo EBC no tiene por qué evitar la inclusión de objetivos de aprendizaje como se ha pretendido hacer, al menos en el caso de México; y (3) la

taxonomía de Marzano y Kendall (2007) ayuda a dar claridad sobre los niveles de procesamiento y dominios de conocimiento que se presentan en el desarrollo de una competencia. Partiendo de estas premisas, el trabajo de investigación que aquí se presenta se realizó a través de construir *a posteriori* los objetivos de aprendizaje con base en la taxonomía de Marzano y Kendall. En otras palabras, partiendo de los contenidos establecidos en el plan de estudios nacional, se generaron un conjunto de objetivos de aprendizaje en que se ubicaban claramente los niveles de procesamiento y dominios de conocimiento que los estudiantes debían demostrar a través de un sistema de evaluación.

Otro referente que fundamenta la investigación que aquí se presenta es el que tuvo que ver con la elección de una teoría psicométrica que permitiera el análisis de los datos que se colectaran. En términos generales existen dos grandes teorías para este fin: la Teoría clásica del puntaje verdadero (Crocker y Algina, 1986) y la Teoría de respuesta al ítem (Hambleton, Swaminathan y Rogers, 1991). De ambas, se optó por la primera para los propósitos de esta investigación. La Teoría clásica sostiene que cada medición del desempeño de una persona se compone de dos elementos: (1) un puntaje verdadero y (2) un error en la medición (aleatorio o no). Esta teoría ha sido fundamental para el desarrollo de pruebas estandarizadas objetivas. Este tipo de pruebas se caracterizan por: (1) permiten evaluar a una gran cantidad de personas en relativamente poco tiempo, bajo condiciones similares; y (2) simplifican el proceso de calificación al no requerir la participación de jueces expertos que interpreten las respuestas de los examinados. Esto trae consigo que se minimicen los errores en el proceso de colección de datos y de calificación de los exámenes, aumentando la confiabilidad de los resultados (Phelps, 2005).

Dentro de las distintas variaciones de pruebas objetivas que hay, para los propósitos de esta investigación, se optó por el uso de preguntas de opción múltiple (Haladyna, 2004). El uso de este tipo de preguntas resulta frecuentemente controversial debido a posturas maniqueas que tienden a ver su uso como restringido a cuestiones meramente memorísticas. Quienes denuestan este tipo de pruebas argumentan que para evaluar desempeño se requiere solicitar al estudiante tareas de producción, más que de selección de una respuesta entre varias opciones. Si bien las preguntas de producción tienen un valor, la crítica que se les hace es que se requieren habilidades concomitantes para responderlas (por ejemplo, la capacidad para redactar), que dificultan distinguir la competencia que se quiere evaluar de la presencia o ausencia de esa habilidad adicional. Las preguntas de opción múltiple, no sólo evitan lo anterior, sino que si están bien planteadas, permiten inclusive evaluar niveles de procesamiento de orden superior, como podían ser los niveles de análisis y utilización del conocimiento (Marzano y Kendall, 2007).

Marco contextual

Dos aspectos que permiten poner en contexto la investigación que aquí se reporta son: (1) algunas características del estado de Tabasco y (2) una breve descripción del Sistema Educativo Mexicano (SEM).

Algunas características del estado de Tabasco. Tabasco es un estado del sureste de México. Tiene por capital a la ciudad de Villahermosa. Está conformado por 17 municipios que integran su división política. El territorio de Tabasco se caracteriza por tener llanuras bajas y húmedas, de origen aluvial, así como depresiones pantanosas inundables. Asimismo, es un estado de una compleja mezcla pluricultural.

Según el censo 2010 realizado en la República Mexicana (ver Instituto de Estadística, Informática y Geografía, INEGI, <http://www3.inegi.org.mx/sistemas/mexicocifras/>), el estado de Tabasco tiene una población de 2,238,603 millones de habitantes, que corresponden al 1.99 % de la población nacional. De esta cantidad, la población económicamente activa fue de 839,766 personas. Según datos de 2008, el Producto Interno Bruto (PIB) del estado de Tabasco fue del 2.58% del PIB nacional. Su principal actividad económica gira alrededor del sector minero, agropecuario, industrial y de servicios.

En el ámbito educativo, según la Prueba ENLACE (ver resultados en el estado de Tabasco, http://www.enlace.sep.gob.mx/gr/docs/historico/27_Tabasco_ENLACE2010.pdf), Tabasco ha estado ubicado en los últimos cinco lugares de la lista nacional. ENLACE es una prueba que permite ubicar a los estudiantes en uno de los cuatro siguientes niveles: insuficiente, elemental, bueno y excelente. Considerando los resultados del nivel primaria (grado 1 a 6), en la materia de matemáticas, se observa que un 87.4% de los estudiantes estuvieron en el 2006 en un nivel de insuficiente y elemental, disminuyendo este porcentaje hasta un 69.2% en el 2010. Para el caso de la materia de español en el mismo nivel educativo, el patrón es similar: un 83.0% estuvo en el nivel insuficiente y elemental en el 2006, mostrando una disminución sistemática hasta llegar al 65.5% en el 2010. A pesar de que se aprecia una mejoría sustancial en un período de cuatro años, estos porcentajes son muy preocupantes ya que hablan de una falta de competencia elemental de los estudiantes para lo que implica el manejo de operaciones básicas de matemáticas y la comprensión lectora. Esta situación es todavía más acentuada en el nivel de secundaria (grado 7 a 9). En la materia de matemáticas, un 97.7% de los estudiantes obtuvo en el 2006 una calificación de insuficiente y elemental, disminuyendo este porcentaje hasta 86.9% en el 2010. En la materia de español el patrón fue similar: el 89.5% de los estudiantes obtuvo una calificación de insuficiente y elemental en el 2006, disminuyendo este porcentaje al 84.0% en el 2010.

Una breve descripción del Sistema Educativo Mexicano (SEM). Además de la información relacionada con las características propias del territorio tabasqueño, es importante visualizar algunas características del SEM. En el momento de realizar esta investigación México iniciaba con un proceso de reforma denominada Reforma Integral de la Educación Básica (RIEB, <http://basica.sep.gob.mx/reformaintegral/sitio>) la cual pretende promover el desarrollo de un conjunto de competencias para la vida de acuerdo con estándares internacionales (e.g., Tuning Project, 2007). El proceso de implementación de la RIEB ha sido gradual: dio inició en el 2006 y terminó en el 2012. No todos los grados académicos estaban trabajando bajo el modelo de competencias al momento en que se realizó la presente investigación. Por esta razón, en primaria específicamente, se tuvieron que revisar los planes de estudio y libros de textos anteriores a la RIEB. Este no fue el caso de secundaria, que en todos los grados escolares la RIEB había sido implementada en su totalidad.

Relevante para esta investigación es hablar sobre la naturaleza de los planes de estudio y libros de texto oficiales. A pesar de que en México, el SEM de escuelas públicas se encuentra descentralizado para fines administrativos, una gran cantidad de decisiones de carácter académico se toman de manera centralizada, a nivel federal. La Secretaría de Educación Pública (SEP) es la entidad que dicta la naturaleza de los planes de estudio y los contenidos de los libros de texto. Asimismo, corresponde a la SEP dictar las políticas que rigen a la RIEB. Por décadas, los libros de texto se han distribuido gratuitamente a estudiantes y maestros. El inicio de esta iniciativa data del año 1950. Actualmente se publican libros para todos los niveles de educación básica, la cual comprende educación preescolar, primaria y secundaria. Adicionalmente, se elaboran también materiales para programas de telesecundaria y de educación indígena. Estos libros integran actividades diseñadas para fomentar los procesos de formación alineados con el modelo educativo vigente en la nación (Limón, 2010). No obstante, el proceso de conformación del material varía para la escuela primaria y secundaria. Por un lado, la emisión de los libros para primaria, es responsabilidad exclusiva de la Secretaría de Educación Pública; mientras que para los de secundaria, se somete a concurso su elaboración, lo cual favorece la inclusión de ideas plurales.

Cabe destacar que el esfuerzo por estandarizar planes y programas de estudio a nivel nacional no excluye la intención de integrar aspectos de las diferentes culturas y costumbres regionales que existen en el país. Es por esta razón que cada entidad federativa desarrolla libros de texto relacionados con geografía e historia regional; así como diversos materiales en lenguas distintas al español (náhuatl, maya, etc.). En los últimos años, las políticas educativas apuntan a la integración escolar de alumnos con capacidades diferentes, razón por la que también se publican libros en sistema Braille para apoyar a los alumnos con limitaciones visuales.

Metodología

Este artículo se refiere a un proyecto titulado: *Evaluar para mejorar: Sistema de evaluación educativa externa*. Como se comentó anteriormente, el proyecto fue dividido en tres etapas: (1) el diseño de instrumentos, (2) la aplicación de éstos en escuelas de educación básica y (3) la capacitación a maestros en diferentes aspectos relacionados con el proceso de enseñanza-aprendizaje. Para los propósitos de este artículo se reportan los resultados de la primera y segunda etapa. Si bien estos resultados tienen relevancia para las autoridades educativas de la entidad, para los interesados en investigación educativa en este artículo se pone énfasis sobre dos aspectos procedimentales: el primero, sobre el uso de la Taxonomía de Marzano y Kendall para el diseño de los instrumentos; y el segundo, sobre la estrategia de involucrar a un equipo de profesores en ejercicio, que muchas veces son soslayados en los procesos de construcción de instrumentos.

Primera etapa: el diseño de los instrumentos. Cuando surge la convocatoria para realizar un estudio de investigación evaluativa en Tabasco, los autores de este artículo enviaron una propuesta tratando de responder a las necesidades planteadas. En dicha propuesta se establecía la evaluación de cuatro disciplinas: español, matemáticas, ciencias naturales y ciencias sociales, por ser consideradas éstas las materias fundamentales en educación básica. Estas disciplinas debían ser evaluadas en los grados cuarto, quinto y sexto de primaria, y primero y tercero de secundaria (en México, la primaria corresponde a los grados 1 a 6, y la secundaria a los grados 7 a 9). En total, 20 instrumentos fueron desarrollados en este estudio. El diseño de los instrumentos se llevó a cabo de acuerdo con los siguientes puntos:

1. *Revisión de planes de estudio y libros de texto.* Para poder determinar los contenidos por evaluar, se revisaron, en un primer término, los planes de estudios y los libros de texto de las cuatro disciplinas antes mencionadas. Aunque los instrumentos de evaluación se aplicarían a los estudiantes de cuarto, quinto y sexto de primaria; y primero y tercero de secundaria, los contenidos que se evaluaron correspondían a los grados inmediatos inferiores.
2. *Identificación de contenidos clave por evaluar.* Doce profesores de primaria y ocho de secundaria, de diferentes partes de la República Mexicana, fueron invitados a participar en la determinación de los contenidos por evaluar, previo sondeo del reconocimiento que otros profesores les daban por su experiencia. La información de planes de estudios y libros de texto fue transcrita en 20 tablas concentradoras. A partir de dicha información, los profesores jerarquizaron la importancia relativa de cada uno de los temas en función del tiempo que los alumnos deben dedicar a su estudio y de la relevancia que tienen para el estudio de temas subsecuentes.

3. *Redacción de objetivos de aprendizaje.* A partir de los contenidos clave seleccionados en el paso anterior, se procedió a redactar un conjunto de objetivos de aprendizaje con base en la Nueva Taxonomía de Marzano y Kendall (2007). Estos objetivos permitieron identificar, para cada contenido por evaluar, los niveles de procesamiento y dominios del conocimiento requeridos por los alumnos. Este proceso es contrario a la lógica natural de un diseño curricular, en el que primero se establecen los objetivos o metas por lograr (o competencias por desarrollar) y posteriormente se definen los contenidos programáticos. El hecho de generar los objetivos de aprendizaje *a posteriori* (a partir de los contenidos en planes de estudio y libros de texto) fue necesario para el eventual diseño de los instrumentos de medición.
4. *Elaboración de tablas de especificaciones.* Una vez identificados los contenidos clave de cada grado académico y de cada disciplina y generados los objetivos de aprendizaje de acuerdo con la Nueva Taxonomía de Marzano y Kendall (2007), se desarrollaron 20 tablas de especificaciones que servirían para guiar la construcción de los instrumentos. Una tabla de especificaciones muestra en sus renglones los contenidos por evaluar; y en sus columnas, los niveles de procesamiento y dominios de conocimiento. La importancia relativa de cada renglón y columna es establecida mediante porcentajes. Esto permite, determinar la cantidad de preguntas o reactivos a diseñar para cada tema y objetivo.
5. *Capacitación a diseñadores de reactivos.* Un grupo de seis de profesores fueron invitados a participar en el desarrollo de un banco de reactivos que posteriormente se utilizaría para elaborar los diferentes instrumentos de evaluación. Previo a la tarea de diseño de los reactivos, se les impartió un curso de capacitación sobre dos temas: la Nueva Taxonomía de Marzano y Kendall (2007) y la técnica para elaborar reactivos de opción múltiple.
6. *Diseño de un banco de reactivos.* A lo largo de 12 semanas, se diseñó un banco de reactivos para cada disciplina y grado escolar, a partir de las tablas de especificaciones definidas anteriormente. El banco tuvo un total de 900 reactivos.
7. *Capacitación a jueces expertos del estado de Tabasco.* Con el fin de evaluar la calidad de los reactivos, se integró un equipo de ocho profesores oriundos del estado de Tabasco. Cuatro de ellos son profesores de primaria y los otros cuatro de secundaria. Estos profesores fueron invitados a participar como jueces expertos, después de un sondeo orientado a identificar el reconocimiento que otros colegas les daban por su experiencia en cada una de las disciplinas que se iban a evaluar. Previo a la tarea de evaluar los reactivos, se les impartió un curso de capacitación sobre dos temas: la Nueva Taxonomía de Marzano y Kendall (2007) y la técnica para elaborar reactivos de opción múltiple.
8. *Evaluación del banco de reactivos.* Los jueces procedieron a la revisión de cada uno de los reactivos propios de su nivel y disciplina. Los criterios para el proceso de evaluación de cada reactivo fueron: (1) vinculación entre los objetivos de aprendizaje y los contenidos del plan de estudios;

- (2) ubicación de cada reactivo en cuanto a su nivel de procesamiento y dominio de conocimiento; (3) aseguramiento de que la respuesta señalada como correcta fuera realmente correcta; (4) corrección en la redacción de cada reactivo; y (5) aseguramiento de que el lenguaje utilizado en la redacción de los reactivos fuera comprendido en el contexto tabasqueño. Del total de 900 reactivos originalmente diseñados, después de la evaluación quedaron 600 reactivos que fueron utilizados para el diseño de los exámenes que se aplicarían en una prueba piloto.
9. *Construcción de los exámenes para prueba piloto.* A partir del banco de reactivos depurado y de las tablas de especificaciones para cada grado y disciplina, se construyeron 20 exámenes que se aplicarían a los estudiantes en una prueba piloto. Para el caso de primaria, los instrumentos para cuarto grado fueron construidos utilizando 20 reactivos; para quinto grado, 25 reactivos; y para sexto grado, 30 reactivos. Para el caso de secundaria, los instrumentos para primer grado tuvieron 35 reactivos; y para tercer grado, 40 reactivos.
 10. *Realización de un estudio piloto.* Los exámenes se aplicaron en cuatro escuelas, dos primarias y dos secundarias, a un total de 200 alumnos. Previo a la aplicación de los instrumentos, se capacitaron a cuatro cuadrillas, de tres personas cada una, para realizar la aplicación de los exámenes y registrar en bitácoras el proceso de examinación.
 11. *Modificación de los exámenes para su versión final.* El estudio piloto fue clave para evaluar la calidad de los reactivos y para revisar la logística de la aplicación de los instrumentos. Con relación al primer punto, se realizó un análisis de reactivos para determinar los índices de dificultad, índices de discriminación y calidad de los distractores de cada uno de los reactivos; así como los índices de consistencia interna (KR-20). Con relación al segundo punto, se tomó la decisión de reducir el número de reactivos por examen, con la finalidad de evitar cansancio y distracciones por parte de los alumnos. Todo esto permitió generar un nuevo conjunto de exámenes para su aplicación a gran escala. Para el caso de primaria, los instrumentos para los tres grados se conformaron por 20 reactivos y para el caso de secundaria, los instrumentos para los dos grados se conformaron por 30 reactivos.

Segunda etapa: aplicación de los instrumentos en una muestra a gran escala. Una vez generados los instrumentos finales, se procedió a la segunda etapa: su aplicación en una muestra a gran escala. La población de interés en este estudio fueron 1920 escuelas primarias y 649 escuelas secundarias públicas del estado de Tabasco. La muestra se eligió de acuerdo con un procedimiento de muestreo aleatorio estratificado y por racimos. Los estratos se definieron a partir de los resultados de la Prueba ENLACE aplicada a la totalidad de escuelas de la entidad en el año 2009. Esto derivó en la elección de 61 escuelas primarias y 57 escuelas secundarias por estudiar, en el entendido de que estas cantidades incluían un 10% extra de escuelas en caso de que alguna no pudiera visitarse por cuestión geográfica o climatológica. Finalmente, en este estudio se visitaron 53

escuelas primarias y 52 escuelas secundarias, ubicadas en 16 municipios de un total de 17 en que se divide el estado de Tabasco. Cada escuela fue considerada como un racimo, lo cual implicó que se aplicarían los instrumentos a todos sus estudiantes en los grados seleccionados. Los exámenes se aplicaron a un total de 814 estudiantes de cuarto, 807 de quinto y 758 de sexto grado de primaria; y a 1199 estudiantes de primero y 1101 de tercer grado de secundaria. En total participaron 4679 estudiantes.

Para aplicar los instrumentos, se conformaron 10 cuadrillas de tres personas cada una: un líder de cuadrilla, que era un profesor, y dos asistentes de investigación. Cada cuadrilla visitó en promedio 10 escuelas en un periodo de seis semanas. Los miembros de la cuadrilla recibieron una capacitación previa sobre los procedimientos para la aplicación de instrumentos y otros aspectos logísticos del estudio. Cabe señalar que este estudio contempló también la aplicación de un conjunto de cuestionarios para recopilar información sobre variables contextuales que permitieran explicar la varianza en los resultados de la aplicación de los exámenes. Estos cuestionarios fueron aplicados a los directores de las escuelas, a los profesores, a los padres de familia y a los mismos estudiantes; sin embargo, para los propósitos de este artículo, esta información no será abordada.

Resultados

De este estudio se desprenden dos tipos de resultados que tienen relevancia para dos tipos de destinatarios: (1) las diversas autoridades educativas que pueden ver, en este caso particular, los niveles de desempeño académico de los estudiantes participantes; y (2) los investigadores y académicos interesados en los procesos de evaluación del aprendizaje. Para los propósitos de este artículo, el primer punto es instrumental para analizar el segundo punto.

Sobre los niveles de desempeño académico. La Tabla 1 muestra los resultados de las pruebas de español, matemáticas, ciencias naturales y ciencias sociales en cuarto, quinto y sexto de primaria, así como en primero y tercero de secundaria. Los resultados incluyen información sobre los estudiantes evaluados (número de estudiantes, promedio y desviación estándar) y sobre las características psicométricas de cada instrumento de medición (porcentaje de reactivos difíciles, porcentaje de reactivos fáciles, porcentaje de reactivos con bajo índice de discriminación y coeficiente de confiabilidad).

Tabla 1
Resultados de la aplicación de los exámenes en primaria y secundaria durante el trabajo de campo

| Disciplina | Indicadores | Primaria | | | Secundaria | | | |
|---------------------|---------------------|--|--|--------|------------|--------|--------|-------|
| | | 4o. | 5o. | 6o. | 1o. | 3o. | | |
| Español | Desempeño académico | Número de estudiantes evaluados | 814 | 807 | 758 | 1,199 | 1,101 | |
| | | Promedio de las calificaciones del grupo en el examen (escala 0 - 100) | 38.63 | 44.59 | 39.26 | 50.47 | 38.95 | |
| | | Desviación estándar de las calificaciones del grupo en el examen | 14.84 | 16.65 | 13.24 | 16.71 | 11.44 | |
| | Datos psicométricos | Porcentaje de reactivos difíciles | 10.00% | 10.00% | 0.00% | 0.00% | 3.33% | |
| | | Porcentaje de reactivos fáciles | 0.00% | 0.00% | 0.00% | 0.00% | 3.33% | |
| | | Porcentaje de reactivos con bajo índice de discriminación | 10.00% | 10.00% | 10.00% | 10.00% | 30.00% | |
| | | Coefficiente de confiabilidad (KR-20) | 0.54 | 0.63 | 0.43 | 0.76 | 0.48 | |
| | Matemáticas | Desempeño académico | Número de estudiantes evaluados | 814 | 807 | 758 | 1,199 | 1,101 |
| | | | Promedio de las calificaciones del grupo en el examen (escala 0 - 100) | 46.73 | 46.05 | 50.53 | 38.88 | 37.90 |
| | | | Desviación estándar de las calificaciones del grupo en el examen | 19.04 | 19.76 | 14.42 | 14.54 | 15.85 |
| Datos psicométricos | | Porcentaje de reactivos difíciles | 0.00% | 0.00% | 10.00% | 10.00% | 3.33% | |
| | | Porcentaje de reactivos fáciles | 5.00% | 0.00% | 5.00% | 0.00% | 0.00% | |
| | | Porcentaje de reactivos con bajo índice de discriminación | 0.00% | 5.00% | 10.00% | 13.33% | 13.33% | |
| | | Coefficiente de confiabilidad (KR-20) | 0.74 | 0.74 | 0.52 | 0.68 | 0.74 | |
| Ciencias naturales | | Desempeño académico | Número de estudiantes evaluados | 814 | 807 | 758 | 1,199 | 1,101 |
| | | | Promedio de las calificaciones del grupo en el examen (escala 0 - 100) | 43.82 | 46.92 | 40.44 | 53.07 | 35.57 |

| Disciplina | Indicadores | Primaria | | | Secundaria | |
|-------------------|--|------------|-----------|------------|------------|------------|
| | | 4o. | 5o. | 6o. | 1o. | 3o. |
| | Desviación estándar de las calificaciones del grupo en el examen | 17.3 9 | 18.4 1 | 13.1 | 15.9 8 | 13.1 1 |
| | Datos psicométricos | | | | | |
| | Porcentaje de reactivos difíciles | 20.0 0% | 0.00 % | 10.0 0% | 3.33 % | 6.67 % |
| | Porcentaje de reactivos fáciles | 0.00 % | 0.00 % | 0.00 % | 10.0 0% | 0.00 % |
| | Porcentaje de reactivos con bajo índice de discriminación | 15.0 0% | 5.00 % | 20.0 0% | 6.67 % | 20.0 0% |
| | Coefficiente de confiabilidad (KR-20) | 0.68 | 0.69 | 0.42 | 0.74 | 0.71 |
| Ciencias sociales | Desempeño académico | | | | | |
| | Número de estudiantes evaluados | 814 | 807 | 758 | 1,199 | 1,101 |
| | Promedio de las calificaciones del grupo en el examen (escala 0 - 100) | 39.5 2 | 43.8 1 | 37.9 7 | 32.6 0 | 40.6 3 |
| | Desviación estándar de las calificaciones del grupo en el examen | 15.3 9 | 15.3 5 | 16.0 1 | 10.5 9 | 14.5 4 |
| | Datos psicométricos | | | | | |
| | Porcentaje de reactivos difíciles | 0.00 % | 0.00 % | 10.0 0% | 6.67 % | 3.33 % |
| | Porcentaje de reactivos fáciles | 5.00 % | 5.00 % | 0.00 % | 0.00 % | 0.00 % |
| | Porcentaje de reactivos con bajo índice de discriminación | 5.00 % | 0.00 % | 5.00 % | 20.0 0% | 6.67 % |
| | Coefficiente de confiabilidad (KR-20) | 0.57 | 0.57 | 0.61 | 0.39 | 0.67 |

Fuente: datos colectados por los investigadores, en trabajo de campo, de enero a julio de 2010.

Acerca del desempeño académico, destaca de manera generalizada un bajo desempeño de los estudiantes en las cuatro disciplinas. Para el caso de primaria, el promedio menor fue 37.97 y el promedio mayor fue 50.53, en una escala del 0 al 100. Para el caso de secundaria el promedio menor fue 32.60 y el promedio mayor fue 53.07 en la misma escala.

En cuanto a los datos psicométricos, la Tabla 1 permite valorar la calidad de cada uno de los 20 instrumentos que se aplicaron en el estudio. A pesar de que se había realizado un estudio piloto para afinar estos instrumentos, fue hasta su aplicación a gran escala que se obtuvieron algunos indicadores *a posteriori*. Uno de esos indicadores es el que corresponde a los índices de dificultad. Para el caso de primaria, de los 240 reactivos que se aplicaron en los 12 instrumentos, 14 de

ellos (esto es, el 5.83%) fueron considerados difíciles y cuatro de ellos (esto es, el 1.67%) fueron considerados fáciles. Un reactivo se consideró difícil si menos del 20% de los estudiantes evaluados lo respondieron correctamente; mientras que un reactivo se consideró fácil si más del 80% de los estudiantes evaluados lo respondieron correctamente. Para el caso de secundaria, de los 240 reactivos que se aplicaron en los ocho instrumentos, 11 de ellos (esto es, el 4.58%) fueron considerados difíciles y cuatro de ellos (esto es, el 1.67%) fueron considerados fáciles.

Otro de los indicadores psicométricos es el que corresponde a los índices de discriminación, definidos como el grado en que un reactivo permite distinguir a los alumnos de alto versus bajo desempeño académico. Los índices de discriminación se determinaron a través de coeficientes de correlación punto-biserial. Entre más alto sea un índice de discriminación, se considera que el reactivo está mejor diseñado o que discrimina mejor (en términos psicométricos) a los alumnos de un mismo grupo. En este caso se contabilizaron aquellos índices bajos de discriminación; esto es aquellos reactivos cuyo coeficiente de correlación fue menor a 0.15. En el caso de primaria, de los 240 reactivos que se aplicaron en los 12 instrumentos, 19 de ellos (esto es, el 7.92%) tuvieron un bajo índice de discriminación. Para el caso de secundaria, de los 240 reactivos que se aplicaron en los ocho instrumentos, 36 de ellos (esto es, el 15.00%) tuvieron un bajo índice de discriminación.

Otro de los indicadores psicométricos es el coeficiente de confiabilidad, el cual fue evaluado en este caso a través del coeficiente de consistencia interna Kuder-Richardson (KR-20). En el caso de los 12 instrumentos aplicados en el nivel de primaria, los coeficientes KR-20 oscilaron en el rango de 0.42 (el menor) a 0.74 (el mayor). En el caso de los ocho instrumentos aplicados en el nivel de secundaria, los coeficientes KR-20 oscilaron en el rango de 0.39 (el menor) a 0.76 (el mayor). En términos generales, se puede afirmar que los resultados que los estudiantes obtuvieron en ambos niveles, tienen una confiabilidad moderada.

Sobre los procesos de evaluación del aprendizaje. Una segunda parte de estos resultados es la que sirve a investigadores y académicos interesados en los procesos de evaluación del aprendizaje. A lo largo del proyecto, los investigadores registraron en diarios de campo (Spradley, 1980) las distintas experiencias del proceso de evaluación previamente descrito en la metodología. De estos diarios emergen cinco dimensiones que incidieron en el proceso de instrumentación: (1) la dificultad de vencer inercias; (2) los referentes previos como facilitadores de nuevos aprendizajes; (3) la falta de conocimientos previos como obstáculos para el desarrollo de la tarea; (4) fondo y forma en la evaluación de reactivos; y (5) la importancia del contexto.

1. *La dificultad de vencer inercias.* Un estereotipo común que tienen las pruebas estandarizadas con reactivos de opción múltiple en México, es su énfasis en la medición de información detallada que los estudiantes deben

recordar. Esta situación es la que se usa para explicar el bajo desempeño que estudiantes mexicanos tienen en pruebas internacionales, como PISA, o en pruebas nacionales como ENLACE. En el desarrollo del proyecto de investigación que aquí se reporta, se observaron dos situaciones que reafirman esta percepción. Por una parte, se observó que los objetivos de aprendizaje plasmados en algunos libros de texto destacan la importancia de que el estudiante “identifique”, “nombre”, “ubique” y “explique”, entre otras conductas, información puntual. Pocas veces se observa en el diseño de los libros de texto la promoción de niveles de procesamiento relacionados con el análisis (“asociar”, “clasificar”, “detectar errores” e “inferir”) y con la utilización del conocimiento (“tomar decisiones”, “generar soluciones”, “experimentar” e “investigar”). Por otra parte, se observó que los profesores, de hecho, tienden a diseñar reactivos que ponen a prueba la capacidad de recordar del estudiante aun en aquellos casos en que un objetivo de aprendizaje explícitamente indica un nivel de procesamiento de orden superior. El involucrar a profesores en este proyecto requirió el vencer inercias vinculadas con la importancia que se le da a la memoria en materia de evaluación.

2. *Los referentes previos como facilitadores de nuevos aprendizajes.* La educación en México está muy impregnada de la taxonomía de Bloom *et al.* (1956). Esto representó una ventaja al momento de introducir la Nueva Taxonomía de Marzano y Kendall (2007) con los profesores que participaron en el diseño de los reactivos. Los principios constructivistas ponen énfasis en apoyar nuevos aprendizajes sobre la base de conocimientos previos y eso fue lo que se hizo en este proyecto, en las etapas de capacitación de los profesores. Al hablar de los niveles de procesamiento, era frecuente observar esfuerzos de analogía entre las taxonomías de Bloom y la de Marzano y Kendall. No ocurrió lo mismo cuando se habló de los sistemas *self* y metacognitivo, ni cuando se habló de los dominios del conocimiento (información y procedimientos mentales), pues éstos no contaban con un precedente en la taxonomía de Bloom.
3. *La falta de conocimientos previos como obstáculos para el desarrollo de la tarea.* El diseño de reactivos de opción múltiple es una tarea multidisciplinaria, en la que el diseñador debe saber acerca de los contenidos que va a evaluar, de las estrategias para construir este tipo de reactivos, de su vinculación con otros aspectos del proceso de enseñanza-aprendizaje (como sería una taxonomía de objetivos) y una buena competencia para redactar. Si bien muchos de los profesores que participaron en este estudio eran reconocidos por su amplia experiencia docente, se apreció un marcado desconocimiento acerca de las técnicas para el diseño de reactivos de opción múltiple y, más todavía, para su evaluación. Aunado a lo anterior, otra área de oportunidad que se observó en los profesores es la que se refiere a la competencia de redacción. Contrario a lo que podría pensarse, el ser profesor no garantiza el tener la competencia para redactar con la precisión y claridad que exige el diseño de reactivos de opción múltiple.

4. *Fondo y forma en la evaluación de reactivos.* La evaluación de la calidad de los reactivos requiere valorar cuestiones de fondo y forma. En este estudio, las cuestiones de fondo rara vez fueron cuestionadas debido esto, en gran parte, al trabajo de identificación de contenidos clave para cada asignatura, en cada grado académico. Como consecuencia, el mayor número de observaciones que los profesores realizaron al evaluar los reactivos se refirieron principalmente a cuestiones de forma, tales como: maneras de hacer más clara la expresión escrita de cada reactivo, la sustitución de palabras por otras que fueran mejor entendidas en la entidad, etc. Un ejemplo de lo anterior se refiere a la palabra “cometa” de uso común en el español estándar. Cuando se redactó un reactivo de matemáticas para calcular el área de un cometa, el primer autor utilizó la palabra “papalote” (de origen náhuatl y de uso común en varias regiones de México). Esta palabra fue sustituida por “papagayo”, de uso común en Tabasco. Un segundo ejemplo se refiere al uso de la expresión “tienda de la esquina”, la cual fue sustituida por la palabra “miscelánea”.
5. *La importancia del contexto.* No sólo ciertas palabras fueron modificadas en la construcción de los reactivos. Para darle más familiaridad al contenido que se incluyó en los instrumentos, se optó por el uso de materiales propios de la región o la entidad, tales como el reglamento de tránsito de Tabasco y artículos del periódico “Tabasco Hoy”, para tener textos que evaluaran lectura de comprensión.

Discusión de resultados

Tal como se mencionó en el apartado anterior, de este estudio se desprenden dos tipos de resultados que tienen relevancia para dos tipos de destinatarios: (1) las diversas autoridades educativas que pueden ver, en este caso particular, los niveles de desempeño académico de los estudiantes participantes; y (2) los investigadores y académicos interesados en los procesos de evaluación del aprendizaje. La discusión se realiza en este mismo orden de ideas.

Sobre los niveles de desempeño académico. Los resultados obtenidos a partir de la aplicación a gran escala de los instrumentos se pueden analizar desde dos perspectivas: el desempeño de los estudiantes y las características psicométricas de los exámenes.

En relación con el desempeño académico, los resultados obtenidos por los estudiantes en este estudio son similares a aquellos que Tabasco ha obtenido en las aplicaciones de la prueba ENLACE desde sus inicios. En ninguno de los exámenes aplicados, a estudiantes de primaria y secundaria, se dio un promedio aprobatorio. En general, se corrobora que el nivel de desempeño académico en la entidad es bajo respecto a lo que marcan los planes de estudio vigentes.

En relación con las características psicométricas de los exámenes, su principal virtud radica en la validez de contenido, como resultado de la forma en que los

instrumentos fueron diseñados. Sin embargo, a pesar de haber seguido de forma rigurosa un proceso de diseño fundamentado en técnicas reconocidas (Crocker y Algina, 1986; Haladyna, 2004; y Thorndike y Hagen, 1989), los datos psicométricos indican que existe una confiabilidad media que lleva a cuestionar cualquier inferencia que se haga acerca de las calificaciones.

Estos resultados permiten proponer varias líneas de investigación para trabajos futuros. Por una parte, de contar con los recursos suficientes, es deseable realizar el estudio con muestras más grandes que permitan establecer los perfiles particulares de cada tipo de escuela primaria (general, multigrado e indígena) y secundaria (general, técnica y telesecundaria) existente en el estado. El estudio actual ofrece tan solo una visión global de todas las escuelas en conjunto. Por otra parte, también hay mucho trabajo por hacer para continuar mejorando los instrumentos; por ejemplo: (1) indagar sobre su validez de criterio y de constructo, (2) determinar la confiabilidad test-retest de los resultados y (3) explorar la hipótesis de que los valores moderados en los índices de confiabilidad son resultado de conductas azarosas (no tomar en serio el examen, responder aleatoriamente a las preguntas, etc.) de los alumnos al momento de presentar el examen.

Sobre los procesos de evaluación del aprendizaje. Una decisión clave para la realización de este proyecto de investigación fue la de involucrar a profesores en el proceso de desarrollar los instrumentos de evaluación del aprendizaje. La experiencia de trabajar con ellos permitió a los investigadores conocer más de cerca las problemáticas a las que se enfrentan los profesores al diseñar exámenes. Incluso, siendo los profesores participantes personas con amplia experiencia docente, quedaron manifiestas, como áreas de oportunidad, (1) la forma de emplear una taxonomía de objetivos de aprendizaje para el diseño de reactivos; (2) el empleo de nuevas taxonomías, como la de Marzano y Kendall (2007), para incorporar en el proceso educativo los avances de la psicología cognitiva; (3) la aplicación de técnicas para el diseño de reactivos de opción múltiple ; y (4) el correcto manejo del lenguaje escrito en la redacción de los reactivos.

Aunque las problemáticas detectadas en este estudio se refieren, principalmente, al diseño y aplicación de instrumentos de medición; en la investigación salieron a relucir otras áreas de oportunidad relacionadas con el desconocimiento de formas de retroalimentar efectivamente a los estudiantes y métodos para analizar psicométricamente los datos colectados al aplicar los exámenes.

Estos resultados permiten proponer varias líneas de investigación para trabajos futuros. Por una parte, se recomiendan estudios de naturaleza cualitativa sobre las prácticas *de facto* que los docentes emplean al evaluar el aprendizaje. Detrás de la forma en que ellos construyen un instrumento, dan retroalimentación a sus alumnos o toman decisiones sobre sus calificaciones, existe un conjunto de creencias y valores subyacentes que convendría explorar. Por otra parte, se

sugieren estudios de investigación-acción asociados con procesos de formación docente sobre temas de evaluación del aprendizaje. El trabajo que se realizó en este estudio fue tan solo la antesala de lo que podría lograrse al extender los procesos de colaboración entre docentes e investigadores en programas de capacitación más estructurados.

Conclusiones

"No dejes que el árbol te impida ver el bosque", dice un proverbio popular. A pesar de que el estudio realizado corresponde a una muestra relativamente amplia de escuelas y estudiantes, la investigación corresponde apenas a un "árbol" dentro de un "bosque" exuberante. Al estudiar al "árbol", se halló que existe un bajo desempeño de los estudiantes tabasqueños en las cuatro disciplinas evaluadas. Los aprendizajes de este estudio también permitieron apreciar cómo el involucramiento de docentes, en el proceso de construcción de los instrumentos de medición, permitió asegurar cierta validez de contenido, aunque la confiabilidad de los resultados fue tan solo moderada.

Este estudio se inserta en el contexto de un "bosque" mucho más grande y complejo. La historia de la evaluación estandarizada en México es reciente (Vidal, 2009). Por ello, no es de extrañar que exista una cultura incipiente de evaluación que está impregnada de falsas creencias y temores por parte de los actores educativos. La promoción de una auténtica cultura de evaluación debe llevar a dichos actores a un conocimiento científico de los principios de evaluación y a un mejor entendimiento del valor que tiene la evaluación dentro de los procesos educativos (Valenzuela, Ramírez y Alfaro, 2010).

La cultura de evaluación depende significativamente del contexto en que la evaluación se lleva a cabo. En México, desde el año 2006, el contexto educativo se ha modificado a la luz de la Reforma Integral de Educación Básica (RIEB), una reforma que enfatiza el trabajo escolar en el marco de un modelo educativo basado en competencias. Dentro de sus fortalezas, la RIEB se caracteriza por sus avances en diseño curricular y de materiales didácticos, incluidos los libros de texto. Sin embargo, una de sus principales debilidades se encuentra en el tema de evaluación; la enseñanza se ha alineado a dicha reforma, pero la evaluación se sigue llevando de una manera tradicional, con énfasis en la evaluación de aspectos memorísticos (Gallardo y Gil, 2011). Aunado a esto, la RIEB parece adolecer de marcos teóricos propios de una educación basada en competencias, como aquéllos que provee la psicología cognitiva a través de la Teoría de novatos-expertos (Chi, Glasser y Farr, 1988 y Gallardo, 2007). Una de las contribuciones principales de este estudio es la propuesta de la Nueva Taxonomía de Marzano y Kendall (2007) como una forma de acercar el concepto de objetivo de aprendizaje a modelos de educación basada en competencias.

Referencias

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of the educational goals. Handbook I: Cognitive domain*. New York, EE.UU.: David McKay.

Chi, M. T. H., Glasser, R. & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ, EE.UU.: Erlbaum.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL, EE.UU.: Holt, Rinehart and Winston.

Gallardo, K. E. (2007). *Educación basada en competencias: propuesta de un modelo de evaluación con base en la teoría de expertos y novatos y aplicado al aprendizaje de Excel* (Tesis doctoral). Tecnológico de Monterrey, Monterrey, NL, México.

Gallardo, K. E. & Gil, M. E. (2011, junio). *Incorporación de la Nueva Taxonomía como referente para el diseño de herramientas de evaluación del aprendizaje conducida en una materia de posgrado en entornos virtuales*. Ponencia presentada en el IX Congreso Internacional sobre Investigación Educativa, Monterrey, NL, México.

Haladyna, T. (2004). *Developing and validating multiple-choice test items* (3a. ed.). Mahwah, NJ, EE.UU.: Erlbaum.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA, EE.UU.: Sage.

Limón, M. A. (2010). ¿Ahora qué sigue? Revalorando el libro de texto gratuito. *Revista AZ*, 30, 12-13. Recuperado de <http://www.revistaaz.com/AZFEBRERO2010.pdf>

Marzano, R. J. & Kendall, J. S. (2007). *The new taxonomy of educational objectives*. Thousand Oaks, CA, EE.UU.: Corwin Press.

Marzano, R. J. & Kendall, J. S. (2008). *Designing and assessing educational objectives. Applying the New Taxonomy*. Thousand Oaks, CA, EE.UU.: Corwin Press.

Phelps, R. P. (Ed.). (2005). *Defending standardized testing*. Mahwah, NJ, EE.UU.: Erlbaum.

Ruiz, M. (2009, julio 6). Herramientas metodológicas en la formación basada en competencias. Recuperado de

<http://www.youtube.com/watch?v=z2aluo052pA&feature=autoplay&list=PL8913642669F08D2F&index=2&playnext=2>

Spradley, J. P. (1980). *Participant observation*. Orlando, FL, EE.UU.: Harcourt Brace Jovanovich.

Thorndike, R. L. & Hagen, E. P. (1989). *Medición y evaluación en psicología y educación* (2a. ed.). Distrito Federal, México: Trillas.

Tuning Project. (2007). Reflexiones y perspectivas de la educación superior en América Latina: informe final América Latina 2004-2007. Recuperado de http://tuning.unideusto.org/tuningal/index.php?d=0&option=com_frontpage&Itemid=1&lang=es

Valenzuela, J. R., Ramírez, M. S. & Alfaro, J. A. (2010). Cultura de evaluación en instituciones educativas: comprensión de indicadores, competencias y valores subyacentes. *Perfiles Educativos*, 33(131), 42-63.

Vidal, R. (2009). *¿Enlace, Exani, Excale o PISA?* Centro Nacional de Evaluación para la Educación Superior (CENEVAL), México. Recuperado de http://archivos.ceneval.edu.mx/archivos_portal/3065/Enlace_Exani_Excali_Pisa.pdf

Woolfolk, A. (2009). *Educational psychology* (11a. ed.). Boston, MA, EE.UU.: Pearson.

Zabala, A. (2010, noviembre 24). Desarrollo de competencias y calidad educativa. *Foro Internacional ESCALAE 2010*. Recuperado de <http://educacionadebate.org.mx/2010/11/24/educacion-por-competencias-desafia-inteligencia-mexicana-2/>

Sitios de Internet de organismos citados en el texto

ENLACE. Resultados a nivel nacional. Recuperado de <http://www.enlace.sep.gob.mx>

ENLACE. Resultados en el estado de Tabasco. Recuperado de http://www.enlace.sep.gob.mx/gr/docs/historico/27_Tabasco_ENLACE2010.pdf

Instituto de Estadística, Informática y Geografía (INEGI). Datos sobre el desarrollo económico de Tabasco. Recuperado de <http://www3.inegi.org.mx/sistemas/mexicocifras/>

Reforma Integral de Educación Básica (RIEB). Recuperado de <http://basica.sep.gob.mx/reformaintegral/sitio>

Secretaría de Educación de Tabasco. Recuperado de <http://www.setab.gob.mx/>

Tuning Project. Recuperado de <http://tuning.unideusto.org/>