Revista Mexicana de Investigación Educativa
ISSN: 1405-6666
revista@comie.org.mx
Consejo Mexicano de Investigación Educativa, A.C.
México

Martínez Rizo, Felipe

LA EVALUACIÓN FORMATIVA DEL APRENDIZAJE EN EL AULA EN LA BIBLIOGRAFÍA EN INGLÉS Y FRANCÉS. REVISIÓN DE LITERATURA

Revista Mexicana de Investigación Educativa, vol. 17, núm. 54, julio-septiembre, 2012, pp. 849-875
Consejo Mexicano de Investigación Educativa, A.C.
Distrito Federal, México

Disponible en: http://www.redalyc.org/articulo.oa?id=14023127008

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

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

RMIE, 2012, VOL. 17, NÚM. 54, PP. 849-875 (ISSN: 14056666)

Investigación

LA EVALUACIÓN FORMATIVA DEL APRENDIZAJE EN EL AULA EN LA BIBLIOGRAFÍA EN INGLÉS Y FRANCÉS

Revisión de literatura

FELIPE MARTÍNEZ RIZO

Resumen:
El uso de pruebas en gran escala para evaluar el aprendizaje se ha extendido en el plano internacional y muchos sistemas educativos nacionales. Las autoridades educativas y amplios sectores de la población esperan resultados positivos para la calidad educativa de esos ejercicios, pero el uso inadecuado de ellos trae consigo también efectos negativos que provocan el rechazo de muchos maestros. Una reacción ha sido el creciente interés por el potencial de la evaluación que hacen los maestros en el aula, en especial con el propósito de aportar elementos para que el aprendizaje mejore, la evaluación llamada *formativa*. En este artículo se presenta el desarrollo de las ideas sobre este tipo de evaluación durante las últimas décadas, en la literatura especializada en inglés y francés.

Abstract:
The use of large-scale tests to evaluate learning has spread at the international level and in many national educational systems. Educational authorities and broad sectors of the population expect positive results in terms of the educational quality of these exercises, but their inadequate use also leads to negative effects that cause widespread rejection among teachers. One reaction has been increased interest in the potential of evaluations teachers carry out in the classroom, especially with the purpose of contributing elements for the improvement of learning; such evaluation is known as formative. This article presents the development of ideas on this sort of evaluation during recent decades, found in specialized literature in English and French.

Palabras clave: Evaluación formativa, teorías de la evaluación, métodos de evaluación, calidad de la educación, revisión de literatura, México.
Keywords: formative evaluation, theories of evaluation, methods of evaluation, quality of education, literature review, Mexico.

Felipe Martínez Rizo es investigador del Departamento de Educación de la Universidad Autónoma de Aguascalientes, Av. Universidad núm. 940, Edificio 13, Ciudad Universitaria, 20131, Aguascalientes, Ags. México. CE: felipemartinez.rizo@gmail.com

Este artículo es producto de un proyecto de investigación que se desarrolló con apoyo del Consejo Nacional de Ciencia y Tecnología.

Introducción

La evaluación siempre ha estado presente en la práctica de los maestros y, en el sentido amplio de la noción de *evaluación formativa* (EF), también lo ha estado en todos los casos en que el docente retroalimenta de alguna manera a sus alumnos cuando verifica sus avances. El sentido actual de la expresión, en cambio, es algo relativamente reciente que cobró fuerza a fines del siglo XX y es todavía un tema novedoso. Esta revisión sobre el tema presenta primero la evolución de la noción en la literatura en inglés, en la que apareció inicialmente en forma explícita, y luego su desarrollo en los medios francófonos.

Para seleccionar los textos a incluir se definió el periodo de 1995 a 2011. Se partió de las principales obras de referencia del campo y las revistas especializadas más importantes sobre el tema, así como algunas bases de datos bibliográficos.

Para la literatura en inglés se revisó la tercera edición de la obra de referencia más importante del campo, muy reciente, que incluye un extenso conjunto de artículos sobre evaluación (Peterson, Baker y McGaw, 2010); la cuarta edición del *Handbook of research on teaching*, en especial artículos de Porter, Young y Odden y de L. Shepard (Richardson, 2008: 259-297 y 1066-1101); los textos de Airasian y Abrams y de Gipps y Stobart en el *International Handbook of Educational Evaluation* (Kellaghan y Stufflebeam, 2003: 533-548 y 549-575); la cuarta edición de la principal obra de referencia sobre medición educativa, *Educational Measurement*, en particular el capítulo *Classroom Assessment*, de L. A. Shepard (Brenan, 2006: 623-646); así como dos obras de referencia recientes enfocadas en particular en el tema: *Formative Classroom Assessment: Theory into Practice* (McMillan, 2007) y *Handbook of formative assessment* (Andrade y Cizek, 2010).

Las revistas revisadas a partir de 1995 fueron *Assessment in Education: Principles, Policy, and Practice; Educational Assessment; Educational Evaluation and Policy Analysis; Educational Measurement: Issues and Practice; Educational Researcher; Practical Assessment Research and Evaluation; y Theory into Practice.*

Se utilizó también la serie de informes técnicos del *Center for Research on Educational Standards and Student Testing* (CRESST), de la Universidad de California en Los Ángeles.

Por lo que se refiere a la literatura en francés, se partió de la revisión de la literatura en esa lengua de Allal y Mottier, publicada en la obra del *Centre for Educational Research and Innovation* (2005), y se revisó la colección completa de la revista *Mesure et Évaluation en Éducation*, que recoge la producción en lengua francesa de Canadá, Francia, Bélgica y Suiza desde 1983.

La evaluación formativa en Estados Unidos y Reino Unido

El Consejo de Autoridades Educativas Estatales (Council of Chief State School Officers) de Estados Unidos define la EF como:

> Un proceso utilizado por maestros y alumnos durante la instrucción, que ofrece retroalimentación para ajustar la forma en que se desarrollan la enseñanza y el aprendizaje, con el propósito de mejorar el logro de los objetivos de enseñanza que se quiere alcanzar.

Susan Brookhart (2005, 2007, 2009) propone una definición que distingue cuatro etapas en el desarrollo de la noción que se sintetizan en el cuadro 1.

CUADRO 1

La definición de evaluación formativa: un concepto en expansión

Scriven, 1967	Información sobre procesos de enseñanza y aprendizaje…			
Bloom *et al.*, 1971		…que el maestro puede usar para tomar decisiones instruccionales…		
Sadler, 1989			…y los alumnos para mejorar su propio desempeño…	
Black-Wiliam, Brookhart, Stiggins…				…y que motiva a los alumnos

Fuente: Brookhart, 2009, con adaptaciones del autor.

Combinando elementos de cada una de las etapas identificadas por Brookhart, la evaluación formativa se puede definir como:

> Un proceso mediante el cual se recaba información sobre el proceso de enseñanza aprendizaje, que los maestros pueden usar para tomar decisiones instruccionales y los alumnos para mejorar su propio desempeño, y que es una fuente de motivación para los alumnos (Brookhar2009:1).

Cada etapa ha aportado algo sustantivo a la noción:

- la idea original de Scriven que distingue evaluación al final o en el proceso;
- la aplicación explícita de la noción a la evaluación del aprendizaje, y no sólo del currículo o programas, por Bloom;
- la identificación de los alumnos como destinatarios clave de la información, con Sadler y su planteamiento de los tres elementos básicos del punto de partida, el de llegada y el recorrido entre uno y otro; y
- el impacto de la *revolución cognitiva* y la atención a la dimensión afectiva, con Black y Wiliam, Shepard, Brookhart y Stiggins, entre otros.

El planteamiento inicial de Scriven

En 1967, y refiriéndose a la evaluación de programas educativos, Scriven propuso la idea original, que distingue la información que se usa para mejorar algo que está en proceso, oponiéndola a la que sirve para valorar el resultado final.

La idea clave del trabajo de Scriven era sencilla: si se evalúa una primera versión de un programa educativo cuando todavía es posible introducir cambios con base en esa evaluación, estamos ante un caso de *evaluación formativa*. Cuando se evalúa la versión final de un programa educativo, para decidir si debe continuar o terminar, se está ante un ejemplo de *evaluación sumativa* (Popham, 2008:3).

Pronto otros estudiosos advirtieron la importancia de esa distinción que hoy parece obvia, pero no se había manejado explícitamente antes del trabajo de Scriven. El primero en hacerlo, y el que tuvo la influencia más duradera, fue Benjamín Bloom.

El paso de Bloom y el modelo del *Mastery Learning*

Un año más tarde la idea fue retomada por Benjamín S. Bloom (1968). Poco después apareció la obra que difundió la noción, aplicándola a la evaluación del aprendizaje de los estudiantes, editada por Bloom, con Thomas Hastings y George Madaus (1971).

En esta obra se precisan las diferencias de las evaluaciones que se usan para apoyar decisiones instruccionales, distinguiendo los propósitos formativos y los sumativos, así como los de ubicación y diagnóstico; se presentan técnicas para la evaluación de objetivos cognitivos y afectivos, se describe la complejidad de los nuevos sistemas de evaluación en gran escala y, en una extensa segunda parte, se presentan once capítulos con ideas sobre otros tantos campos especializados.

El trabajo de Bloom y socios añade un elemento importante a la idea de Scriven: que, además de "ofrecer información sobre el proceso de aprendizaje y no sólo sobre resultados finales, lo que aporta la evaluación formativa puede servir a los maestros para que tomen mejores decisiones instruccionales" (Brookhart, 2009:1).

Guskey señala que Bloom constató que muchas veces el contenido de un examen tiene poco que ver con lo visto en clase, que es también lo que el alumno repasa al prepararse para la prueba, lo que lo lleva a la conclusión de que no vale la pena esforzarse y que no conviene fiarse de los profesores.

> Esta experiencia se deriva de la creencia largamente aceptada, pero falsa, de que las evaluaciones deben guardarse en secreto y de que no es justo dejar que los alumnos sepan lo que se incluirá en ellas. Lamentablemente esto lleva al alumno a ver las evaluaciones como adivinanzas, y a creer que su éxito escolar depende en gran medida de qué tanto puede anticipar lo que sus maestros preguntarán en exámenes y pruebas… es difícil que estas sean las lecciones que un maestro responsable quiere que aprendan sus alumnos. En contraste con lo anterior, Bloom recomendaba que los maestros utilizaran sus evaluaciones como fuentes de información o retroalimentación para los alumnos. Para destacar el propósito "informativo" Bloom (1968) recomendaba llamar a estas evaluaciones "formativas", tomando el término de Scriven (1967) que, el año anterior, lo había utilizado para describir los aspectos informativos, más que enjuiciadores, de las evaluaciones de programas (Guskey, 2007:66).

Para Bloom la evaluación debía servir a los maestros para mejorar la calidad de su enseñanza y debía ir seguida por

> [...] instrucción correctiva de alta calidad, que no es simplemente volver a enseñar, lo que típicamente quiere decir repetir las explicaciones originales en voz más alta y más despacio; más bien implica la utilización de enfoques que adapten la enseñanza a las diferencias de estilos de aprendizaje, modalidades y formas de inteligencia de los alumnos (Guskey, 2007:66-67).

A partir de estas ideas, Bloom desarrolló el sistema de enseñanza conocido como *Mastery Learning*, que propone el uso sistemático de la evaluación formativa, al final de periodos cortos de instrucción, seguida por instrucción correctiva adaptada individualmente, con lo cual el autor esperaba una reducción sustancial de las brechas que separan a los alumnos más y menos avanzados.

El uso de evaluaciones formativas frecuentes se complementa con otro elemento clave, la *diferenciación de la enseñanza* según las características y condiciones de cada alumno; *en otras palabras, para que disminuya la variación de los resultados los maestros deben aumentar la variación de su enseñanza*. Al principio esto implica un avance más lento pero, según los defensores del sistema, en el ciclo escolar no es necesario reducir los contenidos a cubrir, ya que:

> [...] el tiempo que se utiliza para instrucción correctiva o enriquecimiento en las primeras unidades trae consigo importantes beneficios que facilitan las cosas más tarde. El tiempo adicional utilizado al principio se recupera en las unidades posteriores dedicando menos a revisiones y avanzando a un ritmo más rápido" (Guskey, 2010:112, 116).

El sistema de Bloom se extendió ampliamente y dio resultados interesantes, como pusieron en evidencia revisiones de literatura que ejemplifican las citas siguientes:

> Pocas intervenciones educativas de cualquier tipo se asocian de manera consistente con efectos en el rendimiento tan grandes como los que produce el *Mastery Learning*... Evaluación tras evaluación los programas que utilizan este enfoque han producido avances impresionantes (Kulik, Kulik y Bangert-Drowns, 1990:292, en Guskey, 2007:75).

El *Mastery Learning* es identificado regularmente como una de las estrategias de enseñanza más efectivas que los maestros pueden utilizar, en cualquier nivel educativo (Walberg, 1984, en Guskey 2007:75).

La década de 1980: cómo llegar a la meta dado el punto de partida

Sadler (1989) da un paso más en el desarrollo de la noción, al añadir que no sólo el maestro puede usar los resultados de la evaluación formativa, sino también los estudiantes y señala que, para que una evaluación sea formativa deberá incluir la identificación precisa de tres elementos:

- El objetivo a alcanzar al final del proceso de enseñanza y aprendizaje de que se trate, o sea el punto de llegada, el aprendizaje esperado.
- La situación en que se encuentra el alumno según la evaluación realizada, o sea el punto de partida de todo nuevo esfuerzo.
- La forma de pasar del segundo al primero, las indicaciones precisas que orientarán al estudiante para que logre alcanzar el aprendizaje esperado.

Wiliam señala que Sadler retomó la noción de retroalimentación de Ramaprasad, del campo de la electrónica, con la precisión de que el rasgo fundamental para considerar que una información constituye retroalimentación es, precisamente, que tenga algún efecto sobre el sistema en que se general. Sadler comenta:

> Un elemento clave de la definición de Ramaprasad es que la información sobre la brecha entre los niveles actuales y los de referencia se podrá considerar retroalimentación sólo si se usa para modificar dicha brecha. Si la información simplemente se registra, se transmite a un tercer actor que no tiene los conocimientos o la capacidad de modificar el resultado o si está codificada de manera demasiado profunda para que conduzca a una acción apropiada (por ejemplo como una calificación sintética dada por el maestro), entonces el ciclo de control no se puede cerrar, y en vez de retroalimentación efectiva se tienen solamente datos sueltos (*dangling data*) (Citado por Wiliam, 2010:19).

La década de 1990: el impacto de la revolución cognitiva

La evaluación formativa se desarrolló en el marco del modelo de *Mastery Learning* de Bloom, con base en los principios neo-conductistas del diseño

instruccional (enseñanza-prueba-retroalimentación o corrección). Por ello importa explorar cómo se transforma "cuando se basa en otras concepciones de la enseñanza y el aprendizaje; cómo se desarrolla si se basa en principios constructivistas o socio-constructivistas o en las ideas sobre la participación en comunidades de práctica, como proponen los teóricos del aprendizaje situado" (Stiggins y Arter, 2002:4).

Desde los inicios del desarrollo de las pruebas en gran escala algunos de sus promotores más lúcidos, como Tyler, señalaban que también ese tipo de evaluaciones debían verse como parte del proceso de enseñanza-aprendizaje, pero prevaleció un enfoque que, en realidad, las manejaba como un elemento adicional que sólo tenía lugar al final del mismo.

Las pruebas en gran escala tradicionales se desarrollaron durante la primera mitad del siglo XX, marcadas por las concepciones psicológicas y pedagógicas de la época, entre las que destacaba el conductismo. Los avances de las concepciones psicométricas se dieron en forma paralela a la llamada *revolución cognitiva*. Estos desarrollos coinciden en rechazar el planteamiento conductista que reduce el campo de estudio de la psicología a los fenómenos directamente observables, para intentar *abrir la caja negra de la mente*, explorando los procesos que tienen lugar en su interior. La revolución cognitiva, dice Lorrie Shepard, fue:

> […] una rebelión contra la psicología de las diferencias individuales y el conductismo, una de cuyas premisas básicas era el centrar la atención en la adquisición de competencias gracias al refuerzo de conductas observables y no en tratar de explicar los procesos mentales subyacentes (Shepard, 2006:627).

En la medida en que se identifican y exploran los procesos mentales se abren horizontes atractivos para la pedagogía y para las metodologías de evaluación del aprendizaje, en especial las que pretendan servir para propósitos *formativos*, esto es para dar elementos que sirvan para que maestros y alumnos modifiquen sus acciones en consecuencia, para alcanzar mejores resultados.

Shepard señala que "sólo recientemente los especialistas en medición comenzaron a prestar atención al contexto del aula para entender mejor las necesidades de los maestros en lo relativo a la preparación para llevar a cabo evaluaciones". Y añade que otros investigadores, Dorr y Bremme, han concluido que:

> […] los maestros piensan con base en razonamientos prácticos y actúan como clínicos, orientando sus actividades evaluadoras a sus tareas cotidianas, como decidir qué enseñar, y cómo hacerlo con alumnos de diversos niveles de desempeño; monitorear el progreso de los alumnos, para saber cómo ajustar la enseñanza en consecuencia; y asignar calificaciones a sus alumnos con base en su desempeño (Cfr. Shepard, 2006:625-626).

Shepard apunta elementos importantes en lo que se refiere al potencial formativo de las evaluaciones, señalando que:

> […] la evaluación no puede promover el aprendizaje si se basa en tareas o preguntas que distraen la atención de los objetivos reales de la enseñanza. Históricamente, las pruebas tradicionales muchas veces orientaban la instrucción en una dirección equivocada, si centraban la atención en lo que es más fácil de medir, en vez de hacerlo en lo que es más importante de aprender (Shepard, 2006:626)

El interés por la evaluación formativa se deriva en parte de la creciente conciencia de las limitaciones de las pruebas en gran escala y se relaciona con avances paralelos debidos a expertos en diversas áreas de contenidos curriculares, que:

> […] comenzaron también a buscar alternativas a las pruebas estandarizadas para su uso en evaluaciones en el contexto del aula, movidos tanto por el rechazo de los efectos de las pruebas utilizadas para rendición de cuentas, como por los profundos cambios en las concepciones del aprendizaje y del manejo adecuado de los contenidos (Shepard, 2006:626).

Muchos maestros no pueden hacer pruebas de mejor calidad a las desarrolladas en gran escala, que sean más adecuadas para retroalimentar su trabajo y el de sus alumnos. Por ello Silver y Kilpatrick sostenían que:

> […] más allá de la práctica prevaleciente según la cual los maestros preparan sus propias pruebas para que se parezcan en forma y contenido a las de opción múltiple externas, debería hacerse un serio esfuerzo para prepararlas más bien para que puedan conducir lecciones de solución de problemas, y para evaluar

> la habilidad y las disposiciones de sus alumnos al respecto en el marco de esas lecciones (Cfr. Shepard, 2006:627).

El calificativo de revolucionario que se aplica a las teorías cognitivas se justifica plenamente, dadas las enormes repercusiones que tiene tanto para la enseñanza como para la evaluación la idea básica de que es posible explorar los procesos que ocurren en la mente de quienes aprenden o enseñan.

En relación con el aprendizaje de habilidades psicomotrices, o el de competencias como la lectura, que tiene un componente psicomotor, hacen sentido los ejercicios reiterados y estímulos de tipo conductista. La repetición y el ensayo error, en cambio, no son el mejor modo de desarrollar competencias cognitivas complejas en matemáticas, ciencias naturales y sociales, cuestiones tecnológicas, de planeación o gestión avanzada, entre otras.

En esos casos son necesarias acciones más complejas para que los aprendices desarrollen (construyan) los nuevos conocimientos, a partir de los previos, en procesos que implican abandonar concepciones equivocadas (*desaprendizaje o deconstrucción*) que dificultan el aprendizaje (*obstáculos epistemológicos*). Para ello no sirve mucho que el maestro indique al estudiante que aún no consigue dominar cierto tema, que lo vuelva a intentar una y otra vez, sino que es necesario explorar los preconceptos erróneos, identificar las etapas del proceso de construcción del conocimiento nuevo (mapas de progreso), ofrecer ejemplos de productos que se acerquen más o menos al esperado, etcétera.

Una enseñanza con esas características será diferente de la tradicional y mucho más rica; la distinción será similar en lo que se refiere al tipo de evaluación a realizar y al tipo de devolución o retroalimentación a ofrecer a los aprendices. Como dice un trabajo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre el tema, "en última instancia, el propósito de la evaluación formativa es conducir a los estudiantes a desarrollar sus propias habilidades de 'aprender a aprender', que a veces se denominan también 'estrategias metacognitivas'" (CERI, 2005:50).

Según McMillan, si se quiere usar la evaluación formativa para la "comprensión profunda" (*deep understanding*), se necesita meta-cognición y autorreflexión.

> La meta-cognición se refiere al monitoreo, la reflexión y la dirección de su propio pensamiento por parte de los estudiantes. Implica monitorear la com-

> prensión, ser consciente de las estrategias que se usan para aprender y reconocer cuando ciertos objetivos de aprendizaje se han alcanzado o necesitan ser revisados y mejorados.
>
> La meta-cognición es similar a la auto regulación, en el sentido de que supone estudiantes autodirigidos (*self-directed*)… que se involucran activamente en su propio aprendizaje y usan habilidades metacognitivas… Utilizan herramientas de pensamiento como repaso, elaboración y organización, administran su tiempo durante el aprendizaje, toman notas y buscan ayuda cuando la necesitan… Los estudiantes aprenden cómo y cuándo solicitar retroalimentación y tienden a detectar errores y a tener habilidades correctivas (McMillan, 2010:47).

Otro ángulo de la cuestión se refiere a la importancia de los aspectos afectivos en la retroalimentación a los alumnos. El trabajo de psicólogos como Crooks (1988) y Natriello (1987) había mostrado el impacto que tiene dar retroalimentación al alumno sobre su nivel de aprendizaje, a partir de las evaluaciones, en una forma o en otra. En esta dirección, trabajos como los de Brookhart, Black y Wiliam (1998) o, más recientemente, Stiggins (2008), subrayan esta dimensión afectiva.

Stiggins señala que, hasta hace poco, los sistemas educativos consideraban normal y aceptable que sólo una parte de los alumnos alcanzara los objetivos de aprendizaje, mientras un número importante no lo conseguía. El papel de la evaluación consistía en distinguir unos y otros en forma consistente, y los criterios para valorar la calidad de las evaluaciones eran su validez y su confiabilidad. Hoy se espera de las escuelas que hagan que todos los alumnos alcancen los niveles de competencia necesarios para vivir en la sociedad del conocimiento, y es necesario reflexionar sobre el papel y las formas apropiadas para evaluar el aprendizaje en este nuevo contexto.

> Las evaluaciones más válidas y confiables del mundo que tengan como efecto hacer que los alumnos abandonen la tarea desesperanzados no pueden ser consideradas productivas, porque hacen más daño que bien... En el pasado, los marcos de referencia para el control de la calidad de las evaluaciones no tomaban en cuenta su impacto en el alumno; la nueva visión de la excelencia en lo relativo a evaluación, en cambio, pone en el centro de la escena este criterio de calidad (Stiggins, 2008:2-3).

En cuanto al impacto emocional de la evaluación sobre el alumno, Stiggins añade:

> Desde los primeros grados, algunos alumnos [...] obtienen altos puntajes en las evaluaciones y altas calificaciones. El efecto emocional es que se ven a sí mismos como capaces de aprender y se sienten confiados [...]
>
> […] otros alumnos obtienen puntajes bajos en las pruebas y calificaciones malas. Esto los lleva a dudar de su capacidad como aprendices. La falta de confianza en sí mismos los priva de las reservas emocionales para correr el riesgo adicional de seguir intentando. El fracaso crónico es difícil de ocultar y se vuelve penoso: mejor ya no intentarlo (Stiggins, 2008:7).

Stiggins vuelve sobre las implicaciones que tiene el que hoy los sistemas educativos aspiren a que todos los alumnos alcancen los estándares establecidos:

> Cuando los que abandonan sin esperanza son los que todavía no alcanzan los estándares, y cuando los educadores tienen que responder ante la sociedad de que todos los estudiantes los alcancen, tenemos un problema serio. Esos alumnos que dejan de esforzarse no dominarán las competencias básicas de lectura, redacción y resolución de problemas matemáticos, y no llegarán a ser personas que sigan aprendiendo a lo largo de su vida. Si la sociedad quiere que todos los alumnos alcancen los estándares, entonces todos los estudiantes deben creer que pueden conseguirlo; todos tienen que tener la confianza suficiente y la disposición necesaria para enfrentar el riesgo de intentarlo. Cualquier otro estado emocional de cualquier alumno es inaceptable (2008:8).

La importancia del cambio de paradigma evaluativo que implica centrar la atención en los alumnos como usuarios privilegiados de los resultados, y en especial teniendo en cuenta el impacto afectivo de las evaluaciones, no se puede exagerar.

> […] los expertos han cometido el error de pensar que los adultos son los usuarios más importantes de la evaluación [...] esta visión pierde de vista que los alumnos pueden ser tomadores de decisiones de aprendizaje más importantes que los adultos [...] los estudiantes pueden hacer que las decisiones de sus maestros sobre la enseñanza sean irrelevantes; tienen el poder de volver ineficaces a los adultos [...] Si un alumno decide que un aprendizaje está fuera de su alcance o

> que el riesgo de fracaso es demasiado grande o amenazador, entonces, hagamos lo que hagamos los adultos, el aprendizaje termina (Stiggins, 2008:8).

Cuestionamientos y avances de la década de 2000

Durante la última década del siglo XX el interés por la evaluación formativa se extendió ampliamente pero, al mismo tiempo, la noción tendió a desvirtuarse al grado de que ha llegado a ser muy ambigua, al referirse a veces a lo que se puede considerar estrictamente formativo, pero otras a cualquier forma de utilizar datos sobre el rendimiento de los alumnos para tomar decisiones sobre la enseñanza (Cfr. McMillan, 2007:1).

En este proceso influyeron los esfuerzos de algunas empresas comerciales que, para promover sus productos, comenzaron a designar con la etiqueta de *formativas* a pruebas tradicionales que presentaban como si estuvieran en línea con las tendencias más actuales en el campo. McMillan dice al respecto:

> En una interesante estrategia de mercado, las compañías que desarrollan pruebas reconocen la importancia de la evaluación formativa, y ahora están promoviendo evaluaciones que llaman "formativas", pero que es más exacto describir como pruebas referidas a ciertos estándares (*benchmark*), que ofrecen mediciones periódicas en lectura y matemáticas para monitorear el avance de los alumnos hacia el logro de lo que cubren las pruebas de alto impacto que se aplican al fin del año escolar (2007:2).

Esas pruebas de medición de avances, monitoreo o intermedias se ofrecen como si fueran formativas aunque hay grandes diferencias; en general "no ofrecen el detalle necesario para introducir correctivos instruccionales apropiados y tienden a interrumpir la enseñanza con poca retroalimentación útil" (McMillan 2007:3). En el mismo sentido Russell precisa:

> Actualmente hay muchas herramientas computarizadas que pueden usarse para obtener información sobre el aprendizaje de los alumnos. Sin embargo si esa información no está bien alineada con los objetivos de aprendizaje y no es recolectada en momentos cercanos a la instrucción… no cumple los requisitos de la definición de evaluación formativa. La aplicación periódica de pruebas para monitorear el avance hacia objetivos de aprendizaje amplios tampoco cae dentro de esa definición; esas pruebas periódicas son más bien evaluaciones intermedias

> (*interim assessments*). Igualmente, si bien el resultado de evaluaciones sumativas puede usarse para identificar temas y habilidades que los alumnos parecen no dominar o que necesitan más desarrollo, el tiempo que transcurre entre el momento en que se recibe la información y la siguiente oportunidad de reforzar el desarrollo de conocimientos y habilidades particulares impide que las pruebas sumativas se utilicen como formas de evaluación formativa (Russell, 2010:125).

Citando a Edwards, Black y Wiliam y Chappuis, Lisa Abrams refiere que 34 estados americanos contaban (hacia 2007) con bases de datos interactivas que podían usarse para retroalimentar las prácticas de enseñanza, pero añade que, si esos sistemas:

> […] no tienen la capacidad de proporcionar retroalimentación correctiva que informe los procesos tanto de maestros como de alumnos, se les debería considerar más bien pruebas "mini-sumativas" o "sumativas para alarma temprana" (*early warning summative*), y no herramientas formativas [...] por la influencia y la mercadotecnia de los programas de *benchmarking* se está redefiniendo el sentido de la expresión "evaluación formativa", que ha llegado a significar "evaluación sumativa frecuente" (Abrams, 2007:93).

Otros dos de los principales promotores de la evaluación formativa afirman que es necesario que la noción se defina de manera precisa, *porque muchos maestros e investigadores parecen haberla entendido mal*; estos autores dicen que algunos maestros creen que utilizar portafolios, en lugar de o además de los resultados de pruebas impuestas externamente, eso constituye evaluación formativa, cuando en realidad "esa práctica nada tiene de formativa a no ser que haya retroalimentación activa que capacite a los estudiantes para modificar y mejorar su trabajo a medida que construyen sus portafolios" (Black y Wiliam, 2004:22).

Por todo ello un grupo de promotores de estas tendencias sugiere un cambio de terminología y proponen la expresión "evaluación para el aprendizaje" (EPA) (*assessment for learning*), en lugar de la ya trillada de *formative assessment*. Una reunión internacional sobre el tema llegó a las siguientes conclusiones.

> […] evaluación formativa es una frase ampliamente utilizada en el discurso educativo en Estados Unidos, Canadá, Nueva Zelanda, el Reino Unido y Europa [...] pero

> las formas en que las palabras se interpretan y reflejan en la política y la práctica educativa revela muchas veces malentendidos en los principios y distorsiones en la práctica que los ideales originales pretenden promover. Algunos de esos malentendidos y desafíos derivan de que las definiciones son ambiguas, pero otros surgen de apropiaciones deliberadas, con intenciones políticas, de principios que han alcanzado un apoyo significativo por parte de los educadores (TICAL, 2009).

Se señala que los principios de Sadler (precisar dónde están los aprendices al inicio, a dónde se quiere que lleguen y como conseguirlo):

> [...] se han malinterpretado a veces como si fueran una exhortación a los maestros para que apliquen pruebas sumativas frecuentemente a sus alumnos, para valorar los niveles que alcanzan en las escalas establecidas nacionalmente para después corregir las fallas y orientarse a alcanzar el siguiente nivel" (TICAL, 2009).

Los mismos investigadores señalaron que, en esos casos:

> […] alcanzar cierto puntaje se vuelve el propósito de la enseñanza, y el aprendizaje real se ve sacrificado en aras del desempeño en la prueba [...] La lógica del grupo es que cuando tiene lugar el aprendizaje verdadero se manifestará en el desempeño, mientras que lo contrario no se sostiene: el desempeño en una prueba por sí mismo no significa necesariamente que haya habido realmente aprendizaje. Se puede enseñar a los aprendices cómo tener buenos puntajes en las pruebas sin que haya mucho aprendizaje real (TICAL, 2009).

Tratando de evitar malentendidos como los mencionados, en el documento que recoge las conclusiones de esos estudiosos se precisa que la expresión "evaluación para el aprendizaje" denota "el proceso de identificar rasgos del aprendizaje a medida que se desarrolla, empleando cualquier proceso formal o informal que ayude a hacerlo, en primer lugar para que el aprendizaje mismo pueda mejorar…". Se subraya que:

> la EPA "busca, analiza y reflexiona sobre información que proviene de los estudiantes, sus maestros y sus pares, como se revela en el diálogo, la observación y la respuesta de los aprendices a tareas y preguntas… Es parte de la enseñanza cotidiana en aula; una buena parte ocurre en tiempo real, pero otra puede derivar de eventos o episodios de evaluación más formal" (TICAL, 2009).

Y se añade: "lo distintivo de la EPA no es el tipo de información o las circunstancias en que se genera, sino su efecto positivo en el aprendiz. Bien integrada en los contextos de enseñanza/aprendizaje, la EPA pone en marcha a los aprendices hacia un aprendizaje más amplio, a lo largo de la vida" (TICAL, 2009).

A partir de lo anterior se propone la siguiente definición:

> La *evaluación para el aprendizaje* es parte de la práctica cotidiana de estudiantes, maestros y pares, que busca, reflexiona y responde a información que proviene del diálogo, la demostración y la observación, en formas que mejoran el aprendizaje que está en proceso (TICAL, 2009).

Una reciente definición muestra un propósito análogo de diferenciación:

> Una práctica que tenga lugar en el aula será formativa en la medida en que la evidencia sobre el desempeño del alumno se produzca, interprete y use por los maestros, los estudiantes o sus pares para tomar decisiones sobre los pasos siguientes en la instrucción que tienen probabilidad de ser mejores, o de estar mejor fundamentadas, que las decisiones que se habrían tomado en ausencia de dicha evidencia (Black y Wiliam, 2009).

Debe quedar claro que no es el uso de algún tipo especial de evaluación lo que la hace formativa, sino el uso que se hace de los datos. Por eso Andrade concluye:

> Toda definición de evaluación formativa debe basarse en su propósito, que deberá incluir informar sobre el aprendizaje de los alumnos a maestros y directivos para orientarlos en la planeación de la enseñanza y retroalimentar a los alumnos sobre su propio avance para ayudarlos a definir cómo cerrar las brechas entre su desempeño y los objetivos establecidos. La esencia de la evaluación formativa es la acción informada (Andrade, 2010:344-351).

La evaluación formativa en el mundo francófono

Los trabajos siguientes se han desarrollado principalmente en la provincia canadiense de Quebec y las regiones francófonas de Suiza y Bélgica.

Tal vez por la influencia cercana del sistema de investigación estadunidense, ha sido en Canadá donde el trabajo de investigación empírica en

evaluación y temas relacionados se ha desarrollado más, muestra de lo cual fue el surgimiento, en Montreal, en noviembre de 1977, de la Association Professionnelle de Mésure en Éducation (APME), tras dos intentos frustrados en el mismo sentido en 1968 y 1972 (Poulin, 1978:4-5). Posteriormente la agrupación adoptó el nombre de Association pour le Développement de la Mésure et l'Évaluation en Éducation (ADMEE).

En 1978 la APME comenzó a editar el boletín *Mésure en Éducation,* cuyo primer número sólo contenía un modesto artículo de seis páginas sobre *estaninas estandarizadas y normalizadas* (Vachon, 1978), además de un texto sobre la historia de la Asociación y los estatutos de la misma. A partir del volumen 5 (1982) la publicación adoptó el nombre de *Mésure et Évaluation en Éducation*, y comenzó a publicar cinco números al año en vez de cuatro.

Con varios cambios en el número de volúmenes y publicaciones al año, se mantiene hasta la fecha y, desde los años 1990, es editada conjuntamente por la Association pour le Développement de la Mésure et l'Évaluation en Éducation (ADMEE-Europe, que agrupa a investigadores de Francia y regiones francófonas de Bélgica y Suiza) y la ADMEE-Canadá, de Quebec, aunque el peso de los trabajos canadienses sigue siendo mayoritario.

Incluso en esta revista, especializada en temas de evaluación, la cantidad de textos que tienen como tema central la evaluación formativa es reducida: sólo 17 en los 30 años transcurridos de 1978 a 2007, ambos inclusive.

En el primer artículo publicado en *Mésure et Évaluation en Éducation* sobre el tema, Linda Allal (1983) menciona el trabajo de Bloom como *perspectiva inicial*, con un enfoque centrado en aspectos metodológicos y técnicos.

Luego menciona una reunión organizada en Ginebra por ella misma con otros investigadores europeos (Allal, Cardinet y Perrenoud, 1979) en relación con la cual señala que se analizaron más bien:

> […] los aspectos conceptuales, en función de perspectivas teóricas provenientes de varios marcos de referencia: análisis psico-sociológicos de las desigualdades del éxito escolar, concepciones neoconductistas y cognitivistas de los procesos de aprendizaje y desarrollos recientes en el campo de la didáctica de las disciplinas escolares. Desde hace algunos años se esboza un tercer enfoque: a partir de un análisis de varios esfuerzos de maestros por poner en práctica la evaluación formativa, se trata de poner en evidencia los procesos de regulación,

> toma de decisiones y comunicación, que caracterizan a la acción pedagógica en su realidad cotidiana (Allal, 1983:37-38).

Esta referencia muestra tanto la influencia inicial, como la línea que tomarían los trabajos francófonos en los años siguientes, que refleja la revista de la ADMEE con un texto de 1983 (Cardinet); otro de 1984 (Weiss); dos de 1985 (Scallon y Huberman *et al.*); dos de 1988 (Bain y Dassa); uno de 1989 (Scallon); cuatro de 1991 (Perrenoud, Gagné y Thouin, Gadbois *et al.* y Cazabon); uno de 1992 (Dassa y Vázquez-Abad); dos de Thouin en 1993 y 1995; uno más de 1994 (Van Nieuwenhoven y Jonnaert); y uno de Richard, Godbout y Picard en 2000.

Llama la atención que de 2001 a 2007 no se encuentren artículos sobre el tema, cuando el interés sobre el mismo era tan fuerte en los medios anglosajones.

Hay dos visiones de conjunto de los trabajos francófonos sobre evaluación formativa, una firmada por Linda Allal y Lucie Mottier López en la publicación que dedicó al tema la Organización para la Cooperación y el Desarrollo Económico (CERI, 2005), y un capítulo similar, más actualizado, de Mottier López, en la obra colectiva compilada por Rebeca Anijovich (2010).

En los dos trabajos se señala como punto de partida el modelo *Mastery Learning* de Bloom, que en francés se tradujo como *Pédagogie de Maitrise*, y en la obra de Anijovich como Pedagogía del Autocontrol o Pedagogía del Dominio.

En el capítulo que aportó a esta última obra, Mottier López señala que, aunque algunos autores formularon diversas reservas frente al modelo de Bloom, se debe reconocer que introdujo un cambio paradigmático, al dejar de relacionar únicamente la evaluación con la certificación de conocimientos o la selección de estudiantes, para verla como un medio para sostener los aprendizajes: "El modelo [...] se dirige a docentes y apunta una transformación de sus prácticas [...] se trata de una verdadera 'revolución' en materia de evaluación de los aprendizajes de los alumnos" (2010:46).

Según la misma autora, al principio los investigadores francófonos centraron la atención en aspectos particulares de *instrumentación* de la evaluación formativa, a partir del *Handbook* publicado por Bloom y colaboradores en 1971. La atención se centraba en la preparación de instrumentos, pruebas y bancos de ítems, así como en la propuesta de métodos

para analizar los tipos de errores que cometían los alumnos al seleccionar diferentes distractores en pruebas de opción múltiple.

Más tarde se desarrolló una reflexión teórica original sobre la evaluación formativa, en particular en tres aspectos: la psicología del aprendizaje, la didáctica de las disciplinas y los abordajes pluridisciplinares (Mottier López, 2010:46-47).

En cuanto a teorías del aprendizaje, la concepción neobehaviorista de Bloom es contrastada por Allal, según Mottier López (2010: 47), "con el abordaje constructivista y cognitivista que lleva a privilegiar modalidades de evaluación formativa centradas en los procesos cognitivos de aprendizaje y no sólo en los rendimientos de los alumnos y las correcciones".

Otros trabajos relacionan la evaluación formativa con las teorías socioculturales de la enseñanza y el aprendizaje, así como con las del aprendizaje situado y con la zona de desarrollo próximo de Vygotsky. Mottier-López ha explorado "los marcos teóricos de la cognición y del aprendizaje 'situados' para conceptualizar la evaluación como una práctica situada que se co-constituye con los alumnos en el contexto social de cada micro-cultura de clase (o comunidad de aprendizaje)" (2010:47-48).

En el campo de la didáctica se analiza la evaluación:

> [...] en tanto componente del sistema didáctico que pone en relación al docente, al educando y el saber que se debe enseñar, estudiando la evaluación en términos de "contrato didáctico" que vincula las expectativas recíprocas del docente y los alumnos en relación con un contenido o con una tarea dada (Mottier López, 2010:48).

Ejemplos de *abordaje pluridisciplinario* de evaluación formativa son aportaciones como las de Perrenoud, para quien es necesario "articular las orientaciones cognitivas, comunicativas y didácticas de la evaluación formativa en un marco más general de la regulación"; las de Cardinet, que considera la evaluación formativa "como un proceso de comunicación lograda entre el docente y el alumno sobre los objetivos, los criterios y las dificultades de aprendizaje", y las de Bonniol y Vial, que "explotan diferentes implicaciones de las teorías cibernéticas, sistémicas y de la complejidad para pensar la evaluación formativa" (Mottier López, 2010:49).

Reconociendo como punto de partida el modelo de *Mastery Learning* de Bloom, los trabajos de lengua francesa lo han ampliado en al menos cuatro direcciones:

- La integración de la evaluación formativa en el proceso de enseñanza-aprendizaje, como parte integral del mismo y no un "acontecimiento específico que tiene lugar después de una fase de enseñanza, lo que implica diversificar los medios de evaluación, que no deben limitarse a pruebas de opción múltiple y tareas para verificar si los alumnos han comprendido el contenido de la lección", sino que incluyen observación de las actividades del alumno e intercambios entre alumnos, entre otros medios (Mottier López, 2010:50).
- La diferenciación pedagógica, a la que se da mucha importancia en la literatura francófona. "Se reconoce que los objetivos de base exigen ser alcanzados por todos los alumnos, pero se ha destacado la cuestión de una posible adaptación de los objetivos con la finalidad de tener más en cuenta las experiencias culturales y los centros de interés personales de los alumnos" (Mottier López, 2010:54). Este punto se relaciona con el relativo a la regulación (cfr. infra), ya que ésta permite identificar diferencias entre los alumnos, las que a su vez hacen necesario diversificar las actividades de instrucción.
- El compromiso del alumno en su propia evaluación. La perspectiva ampliada francófona, a diferencia de Bloom, para quien era el docente el responsable de la evaluación, alienta un compromiso mayor del alumno en la evaluación formativa, distinguiendo tres variantes: la autoevaluación en el sentido estricto, la evaluación mutua entre pares, y la coevaluación que estipula una confrontación de las evaluaciones realizadas por el docente y el alumno (Mottier López, 2010:52-53).
- La noción de regulación, que constituye posiblemente la aportación fundamental de la perspectiva ampliada de la literatura en francés al campo de la evaluación formativa, en el que "la idea de **remediar** las dificultades de aprendizaje (retroalimentación + corrección) es sustituida por la noción más amplia de **regular** el aprendizaje (retroalimentación + adaptación)" (Allal y Mottier López, 2005:245; Mottier López, 2010:51).

Según la autora que se viene siguiendo en este apartado, este cambio central fue originalmente una aportación de Cardinet (1977), que posteriormente fue seguido por la distinción de Allal de tres formas de regulación, que ha sido retomada ampliamente en los trabajos francófonos:

- La *regulación interactiva*, basada en la interacción del estudiante con el otro polo de la actividad de instrucción que es el maestro, con los demás alumnos y/o con materiales que permiten un aprendizaje autorregulado.
- La *regulación retroactiva*, que se hace después de terminar una etapa de instrucción y permite identificar los objetivos alcanzados o no por cada alumno. Puede ser inmediata (v.gr. *on line*) o diferida, y es la propuesta por Bloom.
- La *regulación proactiva* aprovecha varias fuentes de información para preparar nuevas actividades de instrucción que tengan en cuenta las diferencias de los estudiantes, sea para acciones remediales o, más bien, de diferenciación y enriquecimiento (Allal y Mottier López, 2005:245-246).

Otro autor de habla francesa que ha desarrollado en especial el concepto de regulación, aplicado a la evaluación de enfoque formativo, es Philippe Perrenoud, muy conocido por su trabajo sobre la noción de competencias.

Según un artículo de la tercera edición de la *International Encyclopedia of Education*, Perrenoud describe la regulación del aprendizaje como "el concepto clave en que se basa la EPA, y también se relaciona con esta visión de la evaluación la concepción de Vygotsky, con su noción de la zona de desarrollo próximo (ZDP), que es crucial para entender la manera en que él cree que los alumnos *avanzan hacia la autonomía*. Se menciona también que la noción de ZDP se relaciona con la de andamiaje (*scafolding*) para el aprendizaje de Bruner (Marshall, 2010).

Perrenoud (1998:87) dice que la evaluación formativa se vuelve fuente de regulación porque, "para que haya retroalimentación de manera regular, es necesario que haya una clara comprensión de la forma en que funcionan los alumnos y la manera en que incorporan elementos ajenos a sus propios procesos de pensamiento".

Según el mismo autor, "la influencia regulatoria de la evaluación formativa es débil si se reduce a una evaluación aplicada al final de una etapa de instrucción, que destaque lagunas en el conocimiento, errores y comprensiones insuficientes de un tema, que lleven a ciertas actividades remediales" (Perrenoud, 1998:91). Además:

> Las habilidades involucradas en la regulación dependen tanto de la capacidad de observación y diálogo del maestro como de la manera en que comprende los obstáculos cognitivos que hay en la mente del alumno y analiza lo que impide superarlos.
>
> En esta tarea la empatía y la perspicacia son cualidades útiles, que implican cierto grado de descentralización: la capacidad de ponerse en el lugar de la persona que no sabe, no entiende y no ve una solución. Estas habilidades implican una regulación formativa, en el sentido más amplio de la expresión (Perrenoud, 1998:97).

Por su parte, y de nuevo según Mottier López,

> Laveault amplía la conceptualización de la autoevaluación, agregando las regulaciones motivacionales además de las cognitivas y metacognitivas, y observa que las regulaciones pueden entrañar efectos disfuncionales, pueden ser insuficientes (*underegulation*), erróneas (*misregulation)* o incluso excesivas (*overegulation*) (Mottier López, 2010:46-55).
>
> [concluye señalando que] no sin reconocer el rol esencial de la autoevaluación con fines de autorregulación, estos trabajos muestran que es esencial comprender cómo "fracasa" una autoevaluación/regulación y cuáles son los límites de las que "triunfan" (Mottier López, 2010:46-55).

Laveault señala que los conceptos de regulación y autorregulación han contribuido ampliamente a redefinir la evaluación formativa como objeto de estudio en las publicaciones en francés, y que los currículos por competencias que sustituyen a los objetivos de aprendizaje en varios países tienen una clara influencia de las teorías socioconstructivistas francesas, "que subrayan el papel central del alumno y de la interacción social en la construcción del conocimiento", y añade:

> Los nuevos programas de estudio desarrollados en los países de habla francesa enfatizan todos, de una u otra forma, el papel fundamental de la evaluación formativa. Tal énfasis no es nuevo, ya que las prácticas de evaluación formativa han estado asociadas desde hace tiempo con las concepciones que valoran la diferenciación de la enseñanza y la igualdad de oportunidades de aprendizaje para todos los estudiantes. La introducción de programas basados en competencias,

> sin embargo, ha exigido rediseñar los instrumentos de evaluación formativa e incluir métodos adicionales de reunir información sobre el aprendizaje de los alumnos (Laveault, 2010: 433-434).

Este autor advierte que los estudiosos francófonos han hecho mucho en el terreno teórico, "pero han publicado muy poco sobre la verificación del impacto de la evaluación formativa sobre el aprendizaje" y que se necesita mucha más investigación empírica (Laveault, 2010:434).

Conclusión

Esta revisión de literatura ha mostrado el desarrollo de la noción de evaluación formativa, a partir del trabajo de Benjamín Bloom, de manera que en países como Estados Unidos y Reino Unido se ha vuelto de gran actualidad por su potencial para mejorar el aprendizaje. Se ha visto también cómo fue elaborada en el mundo de habla francesa y en el marco de la psicología cognitiva de Piaget o Vygotsky, con nociones como meta-cognición, regulación y autorregulación.

El interés por la evaluación formativa se ha extendido a otros países fuera de los ámbitos anglosajón y francófono pero, en general, sin aportaciones originales de importancia, por lo que esta revisión se limita a los ámbitos mencionados. Las personas que se interesen por esos trabajos podrán comenzar con la revisión de textos como los de Carless (2010) sobre Hong Kong; Bethell (2010) sobre países de Europa Oriental; o Howie, Zimmerman y Draper (2010) en relación con África. El texto de Anijovich (2010), además de información sobre trabajos francófonos, presenta también información sobre estudios latinoamericanos. El informe final de la investigación de la que se deriva este artículo, que se puede solicitar al autor, presenta información más abundante sobre el tema, incluyendo los resultados de trabajos empíricos sobre el impacto potencial de la evaluación formativa.

Referencias

Abrams, Lisa M. (2007). "Implications of high-stakes testing for the use of formative classroom assessment", en McMillan, 2007:43-62.

Airasian, Peter W. y Lisa M. Abrams (2003). "Classroom student evaluation", en Kellaghan y Stufflebeam, 2003:533-548.

Allal, Linda (1983). "Évaluation formative: entre l'intuition et l'instrumentation", M*esure et évaluation en éducation*, vol. 6 (5): 37-57.

Allal, Linda; J. Cardinet y P. Perrenoud (eds.) (1979). *L 'évaluation formative dans un enseignement différencié*, Berna: Lang.

Allal, Linda y Lucie Mottier López (2005). "Formative assessment of learning: A review of publications in French", en CERI, 2005:241-264.

Andrade, Heidi L. (2010). "Summing up and moving forward: Key challenges and future directions for research and development in formative assessment", en Andrade y Cizek, 2010, cap. 20, pp. 344-351.

Andrade, Heidi L. y G. J. Cizek (eds.) (2010). *Handbook of Formative Assessment*, Nueva York: Routledge.

Anijovich, R. (comp.) (2010). *La evaluación significativa*, Buenos Aires: Paidós.

Bain, D. (1988). "L'évaluation formative fait fausse route", *Mesure et évaluation en éducation*, vol. 10 (4):23-32.

Bethell, G. (2010). "Student assessment: Policy & practice in Eastern Europe", en Peterson, Baker y McGaw, 2010, vol. 3:472-478.

Black, Paul y D. Wiliam (1998). "Assessment and classroom learning", *Assessment in Education*, vol. 5, pp. 7-74.

Black, Paul y D. Wiliam (2004). "The formative purpose: Assessment must first promote learning", en Wilson, Mark (ed.), 2004: 20-50.

Black, Paul y D. Wiliam (2009). "Developing a theory of formative assessment", *Educational Assessment, Evaluation and Accountability,* vol. 21, (1): 5-31.

Bloom, Benjamin S. (1968). "Learning for Mastery", *Evaluation Comment* (UCLA-CSIEP), vol. 1 (2):1-12.

Bloom, Benjamin S.; J. T. Hastings; G. F. Madaus *et al.* (1971). *Handbook on formative and summative evaluation of student learning*, Nueva York. McGraw-Hill.

Brennan, Robert L. (ed.) (2006). *Educational Measurement*, 4ª ed., Westport: Praeger.

Brookhart, Susan M. (2005). *Research on formative classroom assessment: State-of-the-Art,* ponencia en la Conferencia Anual de la AERA, Montreal.

Brookhart, Susan M. (2007). "Expanding views about formative classroom assessment: A review of the literature", en McMillan, 2007:43-62.

Brookhart, Susan M. (2009). "Editorial", *Educational Measurement: Issues and Practice*, vol. 28, núm. 1, pp. 1-2.

Cardinet, J. (1977). *Objectifs éducatifs et évaluation individualisée,* 2ª ed., Neuchatel, Institut Romand de Recherches et de Documentation Pédagogiques, Report núm. R77.05.

Cardinet, J. (1983). "Quelques directions de progrès possibles pour l'appréciation du travail des élèves", *Mesure et évaluation en éducation*, vol. 6 (5): 5-35.

Carless, D. (2010). "Classroom assessment in policy context (Hong Kong)", en Peterson, Baker y McGaw, 2010, vol. 3, pp. 438-442.

Cazabon, B. (1991). "L'évaluation formative de la communication: l'intégration des composantes", *Mesure et évaluation en éducation,* vol. 14 (3): 5-21.

CERI (2005). *Formative assessment. Improving learning in secondary classrooms*, París: OECD-Centre for Educational Research and Innovation.

Crooks, T. J. (1988). "The impact of classroom evaluation practices on students", *Review of Educational Research*, vol. 58, pp. 438-481.

Dassa, Clement (1988). "L'intégration du diagnostique pédagogique aux apprentissages scolaires : de la théorie a la voie informatique", *Mesure et évaluation en éducation,* vol. 11 (1): 7-26.

Dassa, C. y J. Vazquez-Abad (1992). "De l'évaluation informatisée a l'interven-tion pédagogique", *Mesure et évaluation en éducation,* vol. 15 (3): 17-24.

Gadbois, L. *et al.* (1991). "Un instrument de mesure des croyances et attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages des élèves", *Mesure et évaluation en éducation,* vol. 14 (2): 5-24.

Gagne, F. y M. Thouin (1991). "L'évaluation formative des apprentissages en orthographe et attitude des élèves á l'égard de I' évaluation", *Mesure et évaluation en éducation*, vol. 14 (1): 5-15.

Gipps, Caroline y G. Stobart (2003). "Alternative assessment", en Kellaghan y Stufflebeam, 2003:549-575.

Guskey, Thomas R. (2007). "Formative classroom assessment and Benjamin S. Bloom: Theory, research and practice", en McMillan, 2007:63-78.

Guskey, Thomas R. (2010). "Formative assessment: The contributions of Benjamin S. Bloom", en Andrade y Cizek, 2010:106-124.

Howie, S. J.; L. Zimmerman y K. Draper (2010). "Classroom assessment in policy context (Sub-Saharan Africa)", en Peterson, Baker y McGaw, 2010, vol. 3, pp.449-456.

Huberman, M.; P. A. Juge y P. A. Hari (1985). "La pédagogie de maitrise: une évaluation instructive au niveau gymnasial", *Mesure et évaluation en éducation,* vol. 8 (3):43-81.

Kellaghan, TH. y D. Stufflebeam (eds.) (2003). *International Handbook of Educational Evaluation*, Dordrecht: Kluwer Academic Publishers.

Kulik, C. C.; J. A. Kulik y R. L. Bangert-Drowns (1990). "Effectiveness of mastery learning programs: A meta-analysis", *Review of Educational Research,* vol. 60 (2):265-299.

Laveault, D. (2010). "Classroom assessment policy context (French Sources)", en Peterson, Baker y McGaw, 2010, vol. 3, pp. 432-437.

Marshall, B. (2010). "Formative assessment and instructional planning", en Peterson, Baker y McGaw, 2010, vol. 3, pp. 365-368.

McMillan, James H. (ed.) (2007). *Formative Classroom Assessment: Theory into Practice*, Nueva York: Teachers College Press.

McMillan, James H. (2007). "Formative classroom assessment: The key to improving student achievement", en McMillan, 2007: 1-7.

McMillan, James H. (2010). "The practical implications of educational aims and contexts for formative assessment", en Cizek, G. y Andrade, 2010:41-58.

Mottier López, Lucie (2010). "Evaluación formativa de los aprendizajes. Síntesis crítica de los trabajos francófonos", en Anijovich, 2010:43-71.

Natriello, G. (1987). "The impact of evaluation processes on students", *Educational Psychologist*, vol. 22, pp. 155-175.

Perrenoud, Philippe (1991). "Pour une approche pragmatique de l'évaluation formative", *Mesure et évaluation en éducation*, vol. 13 (4):49-81.

Peterson, P.; E. Baker y B. McGaw y (eds.) (2010). *International Encyclopedia of Education,* 3ª ed., Amsterdam: Elsevier-Academic Press.

Perrenoud, Philippe (1998). "From formative evaluation to controlled regulation of learning processes. Towards a wider conceptual field", *Assessment in Education*, vol. 5, núm. 1.

Popham, W. James (2008). *Transformative Assessment*, Alexandria: Association for Supervision and Curriculum Development.

Porter, A.; P. Young y A. Odden (2008). "Advances in teacher assessments & their uses", en Richardson, 2008:259-297.

Poulin, Benoit (1978). "Historique de l'Association", *Mesure et évaluation en éducation,* vol. 1 (1):4-5.

Richard, J. F. ; P. Godbout e Y. Picard (2000). "La validation d'une procédure d'évaluation formative en jeux et sports collectifs", *Mesure et évaluation en éducation*, vol. 23, (1):43-67.

Richardson, Virginia (ed.) (2008). *Handbook of research on teaching,* 4ta. *edición,* Washington: AERA.

Russell, Michael K. (2010). "Technology-aided formative assessment of learning: New developments & applications", en Andrade y Cizek 2010:125-138.

Sadler, D. R. (1989). "Formative assessment and the design of instructional systems", *Instructional Science*, vol. 18, pp. 119-144.

Scallon, G. (1985). "La participation des élèves au diagnostic pédagogique: exploration avec des élèves de 4e secondaire en mathématiques", *Mesure et évaluation en éducation*, vol. 8 (1-2): 5-44.

Scallon, G. (1989). "Plaidoyer pour une méthodologie instrumentée d'évaluation formative", *Mesure et évaluation en éducation*, vol. 11 (1): 43-55.

Scriven, Michael (1967). "The methodology of evaluation", en R. W. Tyler, R. M. Gagne y M. Scriven (eds.).

Shepard, Lorrie A. (2006). "Classroom Assessment", en Brennan, 2006:623-646.

Shepard, Lorrie A. (2008). "The role of classroom assessment in teaching and learning", en Richardson, 2008:1066-1101.

Stiggins, Richard J. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*, Portland: ETS-ATI.

Stiggins, Richard J. y J. Arter, (2002). "Assessment for learning, International perspectives. The proceedings of an international conference" (International Conference on Assessment for Learning, Chester, 2001), paper presented at the Annual Meeting of the NCME, Nueva Orleans, abril.

Thouin, M. (1993). "L'évaluation des apprentissages en mathématiques: une perspective constructiviste", *Mesure et évaluation en éducation,* vol. 16 (1-2): 47-64.

Thouin, M. (1995). "Le développement d'instruments de mesure des apprentissages en sciences de la nature au primaire", *Mesure et évaluation en éducation,* vol. 18 (1): 95-124.

Tyler, R. W.; R. M. Gagne y M. Scriven (eds.) (1967). *Perspectives of Curriculum Evaluation*, Chicago: Rand McNally.

TICAL (2009). *Position paper on assessment for learning*, Third International Conference on Assessment for Learning. Dunedin, Nueva Zelanda, marzo.

Vachon, Jean (1978). "Staneuf standardisés et staneufs normalisés", *Mesure et évaluation en éducation,* vol. 1 (1):19-24.

Van Nieuwenhoven, C. y P. Jonnaert (1994). "Une approche des représentations des enseignants du primaire propos de l'évaluation", *Mesure el évaluation en éducation,* vol. 16 (3-4):41-79.

Walberg, H. J. (1984). "Improving the productivity of America's schools", *Educational Leadership,* vol. 41 (8):19-27.

Weiss, J. (1984). "Heurs et malheurs d'un instrument d'évaluation", *Mesure et évaluation en éducation,* vol. 7 (1): 31-42.

Wiliam, Dylan (2010). "An integrative summary of the research literature and implications for a new theory of formative assessment", en Andrade y Cizek, 2010: 18-40.

Wilson, Mark (Ed.) (2004). *Towards coherence between classroom assessment and accountability,* 103rd Yearbook, part 2, Nat. Soc. for the Study of Education, Chicago: University of Chicago Press.

Artículo recibido: 3 de octubre de 2011
Dictaminado: 11 de enero de 2012
Segunda versión: 30 de enero de 2012
Aceptado: 31 de enero de 2012