

USO FORMATIVO DE LA EVALUACIÓN PARA MEJORAR EL APRENDIZAJE. ESTADO DEL CONOCIMIENTO Y DIAGNÓSTICO DE PRÁCTICAS DE MAESTROS DE PRIMARIA

FELIPE MARTÍNEZ RIZO / GUADALUPE RUIZ CUELLAR

Departamento de Educación, Universidad Autónoma de Aguascalientes

RESUMEN: La ponencia presenta resultados parciales de un proyecto en proceso de realización que tiene dos grandes partes: un estado del conocimiento (terminado) basado en la revisión de una extensa literatura (unas 800 referencias) que cubren la producción sobre el tema a nivel internacional y muestran que se trata de un campo en pleno desarrollo; y un diagnóstico de las prácticas de evaluación de maestros de una muestra representativa de primarias generales públicas de Nuevo León, en las que se aplicaron cuestionarios a cerca de 5,000 alumnos de sexto grado y alrededor de 500 maestros de segundo, quinto y sexto. Además se hicieron entrevistas a un

centenar de esos maestros y se fotografiaron muestras de exámenes aplicados por ellos y tareas asignadas a sus alumnos. Los resultados parciales permiten adelantar ya que las prácticas predominantes no corresponden a los planteamientos de las concepciones actuales sobre la evaluación formativa, y que una modificación de dichas prácticas supone cambios en profundidad tanto de las concepciones de los docentes, como de las condiciones en las que suele desarrollarse su trabajo en las primarias mexicanas.

PALABRAS CLAVE: Educación básica, estado del arte, evaluación formativa, práctica docente.

El estado del conocimiento

Prácticas tradicionales de evaluación

En las escuelas antiguas el número de alumnos era reducido y la evaluación no implicaba procedimientos formales; bastaba el juicio del maestro, basado en preguntas y observación. Todavía en el siglo XIX las concepciones pedagógicas conservaban rasgos de *una forma de instrucción elemental que se desarrolla oralmente, en la forma de preguntas y respuestas, y se conoce como catecismo... al masificarse los sistemas educativos la enseñanza tipo catecismo pasó a las aulas laicas*. Los contenidos se ampliaron pero la forma de interacción permaneció notablemente constante. Después de que los alumnos leían un texto o resolvían ejercicios aritméticos, el profesor hacía preguntas... *La forma del*

catecismo todavía está presente en lo que ahora llamamos “enseñanza estándar basada en recitación”. (Resnick et al., 2010: 400)

En el siglo XX una reacción ante las limitaciones de las evaluaciones tradicionales fue el desarrollo de pruebas estandarizadas, cuyos impulsores desarrollaron instrumentos que permitieran comparar los niveles de alumnos de diferentes escuelas. La comparabilidad era atractiva pero las pruebas también tenían limitaciones. Pese a ello predominó la tendencia a verlas como la forma preferida de evaluación, que los maestros deberían imitar.

En las últimas décadas del siglo XX el uso de pruebas se extendió en el mundo despertando grandes expectativas y provocando reacciones negativas ante la reducción del currículo, la preparación de los alumnos para las pruebas y su uso para evaluar escuelas y maestros sin tener en cuenta que sus resultados no bastan para ello. Otra reacción busca fortalecer la evaluación que hacen los maestros, que puede impactar favorablemente el aprendizaje.

La Evaluación Formativa (EF)

En sentido amplio la EF ha estado presente desde hace mucho en la práctica de todo maestro que retroalimenta a sus alumnos al verificar su avance. En sentido estricto es relativamente reciente. En forma explícita la EF apareció en 1967 con Scriven; Bloom la desarrolló fuertemente con el modelo *Mastery Learning*; en los 90 se enriqueció con los planteamientos de Sadler y después con los derivados de la *revolución cognitiva*. En el marco del *Mastery Learning* se desarrolla *con base en los principios neo-conductistas del diseño instruccional (enseñanza-prueba-retroalimentación o corrección)*, pero se transforma cuando se basa en otras concepciones de la enseñanza y el aprendizaje, como en *principios constructivistas o socio-constructivistas o en las ideas sobre la participación en comunidades de práctica, como proponen los teóricos del aprendizaje situado* (Brookhart, 2007; Stiggins y Arter, 2002: 4).

Una enseñanza que siga los principios derivados de la revolución cognitiva es diferente de la tradicional, es más rica y lo mismo ocurre en cuanto a evaluación y retroalimentación: *En última instancia el propósito de la EF es conducir al alumno a desarrollar sus habilidades de “aprender a aprender”, sus “estrategias metacognitivas”* (CERI, 2005: 50).

A fines del siglo XX el interés por la EF se extendió, pero la noción tendió a desvirtuarse: las compañías que desarrollan pruebas comenzaron a promover evaluaciones que llama-

ban formativas pero en realidad eran mediciones periódicas en lectura y matemáticas para monitorear el avance en relación con las que se aplican al fin del año escolar. Esas pruebas realmente no son formativas, pues no dan el detalle necesario para introducir correctivos instruccionales y tienden a interrumpir la enseñanza con poca retroalimentación útil.

La evaluación en México

La evaluación tradicional en México presentaba también el esquema *Iniciación-Respuesta-Evaluación* derivado del catecismo. A mediados del siglo XX, desde antes de su uso en primaria, las pruebas en gran escala influían en la evaluación de los maestros. Comenzó a pedirse a los maestros que hicieran pruebas de opción múltiple y estandarizaran el puntaje de sus alumnos, lo que se llamaba calificar por curva. En paralelo surgieron dos prácticas:

- ◆ Reducción de reprobación y repetición sin medidas para asegurar logro de objetivos por todos los alumnos, buscando sólo reducir acumulación de alumnos en primeros grados.
- ◆ Proliferación de pruebas de mala calidad ofrecidas por supervisores y editoriales, atendiendo demanda de muchos docentes presionados por autoridades y sin preparación para hacer ellos mismos buenas pruebas.

La literatura especializada mexicana sobre EF es casi inexistente; las pocas orientaciones al respecto se encuentran en acuerdos secretariales que mencionan importancia de evaluación pero dan directrices incongruentes con EF y se limitan a precisar aspectos administrativos.

Los materiales del currículo derivado de la Reforma Integral de Educación Básica (RIEB) 2009 son más precisos.

El Módulo 3 del *Diplomado para Docentes de Primaria* se dedica al tema de la *Evaluación para el aprendizaje en el aula*. Discute la importancia de transitar hacia EF que favorezca el proceso de enseñanza aprendizaje y dice que *...esta visión de la evaluación implica también un cambio radical en nuestra sociedad, significa el poder observar en los alumnos la movilización de saberes en situaciones concretas de la vida cotidiana, identificar qué saben hacer con conocimiento y comprometerse en apoyarlos en lo que aún les falta*

por lograr. Se propone que los docentes analicen las prácticas escolares de evaluación generalizadas en la escuela y perciban el potencial educativo de la evaluación para promover el aprendizaje de los alumnos... (Díaz Barriga *et al.*, 2010: 10).

Los propósitos son positivos, pero se aprecian puntos débiles. Hay una visión esquemática; la terminología es novedosa pero no se tiene en cuenta la dificultad de la EF, si se quiere cambiar en profundidad las prácticas de los docentes. De la bibliografía anglosajona solo hay una referencia a Scriven, pionero pero con planteamientos elementales e incluso parece que se le conoce superficialmente. Se toman ideas de referencias españolas y francófonas pero no se aprovechan las ricas aportaciones cognitivistas.

Los pocos estudios empíricos mexicanos sobre EF en la última década muestran panorama que sigue presentando combinación de prácticas tradicionales con evaluaciones inspiradas en las pruebas del siglo XX y poco de los nuevos enfoques.

Estudios empíricos sobre el impacto en el aprendizaje

Textos muy difundidos (v. gr. Black-William 1998) han fomentado idea optimista excesiva de la EF. Una lectura atenta de esos textos no deja de advertir limitaciones tan claras que sorprende que no sean atendidas en toda su importancia. En este sentido destaca una cuidadosa crítica de Black y William que dice:

Una creencia casi nunca cuestionada es que la investigación demuestra en forma concluyente que el uso de evaluación formativa facilita la mejora de las prácticas de enseñanza, identifica lagunas en el currículo y contribuye a aumentar el desempeño de los alumnos. Sin embargo... una revisión de la literatura reveló la limitada evidencia empírica que demuestra que el uso de evaluación formativa en el aula resulta directamente en cambios marcados en los resultados educativos (Dunn y Mulvenon, 2009: 1).

La conclusión del trabajo precisa que:

En cierta medida, las investigaciones discutidas... apoyan el impacto de la evaluación formativa sobre el rendimiento de los alumnos, pero en una medida mayor apoyan la necesidad de hacer investigaciones en las que diseños y metodologías más eficientes lleven a resultados más concluyentes... no argumentamos que la evaluación formativa carezca de importancia, sino sólo que la evidencia empírica que existe para apoyar "las mejores prácticas" de EF es limitada (Dunn y Mulvenon, 2009: 9).

Otros trabajos recientes utilizan acercamientos metodológicos más sólidos. Unos se refieren a experiencias de amplia duración, que involucraron a números importantes de maestros, alumnos y grupos y que, a falta de diseños experimentales estrictos usaron acercamientos cuasi-experimentales cuidando la comparabilidad de los grupos con y sin tratamiento y utilizando técnicas analíticas avanzadas.

Un trabajo concluye con comentarios sobre retos metodológicos incluyendo la dificultad de manejar diseños experimentales con grupos completos, las de la medición de resultados, la muestra y la fidelidad de implementación. En muchos casos los resultados apoyan la hipótesis de que las prácticas de EF contribuyen a mejorar el aprendizaje; un número menor no encontró diferencias significativas. La evidencia no es concluyente pero inclina la balanza en sentido de las opiniones favorables a la EF (Schneider-Randel, 2010: 267-272).

Otro trabajo buscó corroborar la afirmación de Black y Wiliam de que la EF tiene impacto positivo en el aprendizaje poniendo a prueba una gran idea: *que se podría obtener una gran ganancia con una inversión relativamente pequeña: incorporar a un currículo de ciencias utilizado en todo el país evaluaciones formativas conceptualmente coherentes.*

Después de describir el estudio, cuidadosamente diseñado e implementado, se reportan los resultados relativos a los cambios esperados en el rendimiento de los alumnos como sigue:

Sorprendentemente los resultados no corroboraron la hipótesis... el grupo experimental no obtuvo resultados mejores que el de comparación ni en las pruebas de rendimiento ni en las medidas de motivación. De hecho los alumnos del grupo de comparación tuvieron resultados promedio ligeramente mejores que los del grupo experimental, aunque no estadísticamente significativos... la brecha entre los alumnos de alto y bajo rendimiento en el grupo experimental no fue tan grande como en el grupo de comparación... (Ruiz Primo et al., 2010: 151).

La revisión de videos de clases grabados durante el estudio permitió buscar una explicación de los resultados revisando la fidelidad de implementación. La conclusión fue que hubo fuertes diferencias en la forma de hacer las evaluaciones, lo cual se refleja en un impacto diferencial en el rendimiento:

...con base en la evidencia recogida durante el estudio de implementación sabemos que, en general, los maestros conseguían que los alumnos compartieran sus ideas, pero que no

conseguían utilizar la información para ajustar su propia enseñanza. Obviamente recomendar que se ajuste la enseñanza y se actúe es más fácil que hacerlo (Ruiz Primo et al., 2010: 154).

También afecta el impacto de la EF la dificultad de aceptar prácticas que confrontan tradiciones negativas pero muy arraigadas, asociadas a las formas tradicionales de evaluar que prevalecen en las aulas. Otro trabajo que tiene que ver con esa resistencia reporta resultados sobre las reacciones de alumnos que participaban en un proyecto de EF, que incluía dejar de dar calificaciones, para desalentar la tendencia a trabajar en función de la nota, y no del interés intrínseco por el aprendizaje.

Cuando se preguntó a esos alumnos cómo se sentían:

...un número considerable tenían opiniones muy negativas, particularmente porque, en su opinión, el hecho de no recibir calificaciones no les permitía saber cómo orientar sus esfuerzos... cuando se les preguntaba si los comentarios que recibían eran útiles, la mayoría opinaba que no les daban suficiente información para saber cómo mejorar. Tampoco pensaban que el hecho de dar calificaciones estigmatizaría a los de bajo rendimiento... el deseo de recibir calificaciones era tan fuerte que algunos admitían que intentaban calcularlas, sobre todo en matemáticas y lengua, en las que los chicos admitían que sumaban las palabras correctamente deletreadas (correct spellings) en las pruebas de vocabulario para calcular cual sería la calificación que habrían recibido (Smith y Gorard, 2005: 31-33).

Un estudio de implementación diría que la puesta en práctica los principios de la EF no fue adecuada, y es que la aplicación en aula de los principios no es sencilla, en particular si se trata de habilidades cognitivas complejas y no de simples tareas memorísticas, ya que para ello no basta que se modifiquen las prácticas de evaluación, sino que es todo el enfoque de la enseñanza lo que debe cambiar.

El diagnóstico de las prácticas de maestros de Nuevo León

Estrategias de obtención de información

El diagnóstico centra la atención en los maestros ya que el objetivo es obtener una visión de la forma en que llevan a cabo la evaluación de los aprendizajes de sus alumnos a lo largo del ciclo escolar. Se tomó la decisión de incluir sólo a los maestros de segundo, quinto y sexto grados de primaria, ya que en ellos la Reforma Integral de la Educación Básica estaba en aplicación al momento del trabajo de campo, y esto tiene implicaciones

relevantes para el tema del proyecto dado que uno de los elementos atendidos en las actividades de actualización relacionadas con dicha Reforma es el relativo a la evaluación que realizan los maestros.

Para observar la práctica docente tradicionalmente se han usado cuestionarios simples y observación. Los primeros son baratos, pero dan información pobre; la segunda da información rica pero requieren tiempo y recursos humanos caros. Por ello se han buscado acercamientos que den buena información con costos razonables, como autorreportes detallados y situados, en la forma de diarios o bitácoras; cuestionarios con preguntas detalladas y contextualizadas (viñetas); portafolios de evidencias, incluyendo análisis de tareas y otros materiales; y acercamientos basados en el uso de grabaciones en video.

Con la idea de aprovechar las ventajas de diferentes acercamientos, el diagnóstico realizado en este estudio combina varios instrumentos: un cuestionario estructurado aplicado a todos los maestros de la muestra de escuelas primarias de los grados segundo, quinto y sexto; una entrevista semi-estructurada aplicada a un maestro por escuela, y la recolección de evidencias de sus prácticas de evaluación en Español y Matemáticas. El cuestionario permite una aproximación gruesa al tema mientras que los otros instrumentos favorecen una indagación en profundidad de los aspectos más finos de las concepciones y prácticas de evaluación en aula. Además, se aplicó un cuestionario estructurado a alumnos de sexto grado para obtener información complementaria sobre algunos aspectos de las prácticas de evaluación de los docentes.

Selección de la muestra

El diagnóstico realizado es representativo de las escuelas primarias públicas de organización completa del estado de Nuevo León. Se decidió no incluir en la muestra primarias indígenas ni cursos comunitarios porque la primera de estas dos modalidades no está presente en la entidad y porque el número de los segundos es muy reducido. Tampoco se incluyen primarias de sostenimiento privado, porque sus condiciones son diferentes de las públicas, que son el foco de la atención de este trabajo.

El esquema de muestreo aplicado es el estratificado por conglomerados en una sola etapa, es decir, se seleccionaron escuelas y dentro de ellas se entrevistó a todos los docentes de segundo, quinto y sexto grados. Para mejorar las estimaciones obtenidas a partir de la muestra fue necesario estratificar en función del tipo de la localidad en la cual se

encuentra asentada la escuela. De este modo, usando la clasificación Urbano-Rural que define Conapo se estratificaron las escuelas en rurales y urbanas. Esta clasificación tiene la ventaja de que está disponible para todas las escuelas, además de tener una gran correspondencia con el tipo de localidad definida en función del número de habitantes, que es la clasificación que habitualmente se utiliza para este tipo de estudios. El tamaño de muestra se obtuvo para una confianza del 95%, un error relativo del 15% y una tasa de respuesta de al menos el 90%. Además se aplicó factor de corrección por finitud para la obtención del tamaño de muestra. De acuerdo a los criterios especificados se obtuvieron las siguientes cifras para la muestra: total de escuelas en muestra = 101; total de docentes = 480. En cuanto a alumnos, se tomó la decisión de aplicar a todos los de sexto grado lo que da un total aproximado de 4,500.

Trabajo de campo

El trabajo de campo estuvo a cargo de estudiantes de sexto semestre de la Lic. en Educación Primaria de la Escuela Normal Prof. Serafín Peña, de Montemorelos, Nuevo León, coordinados por los profesores del cuerpo académico sobre Evaluación de Aprendizajes de dicha institución y fue realizado en la primera quincena de abril de 2011. Previamente, el equipo responsable del proyecto capacitó al grupo de aplicadores y entregó todos los materiales necesarios para efectuar el levantamiento de información: un Manual para la Aplicación, los cuestionarios para alumnos y maestros, la guía para la entrevista a un maestro por escuela, indicaciones para los propios maestros sobre la recolección de evidencias de sus prácticas de evaluación, cartas de presentación y agradecimiento y el equipo necesario para la recolección fotográfica de exámenes aplicados y tareas solicitadas a los alumnos para evaluarlos en las materias de Español y Matemáticas.

En cada escuela el trabajo fue realizado en una jornada. Al inicio de esta, el aplicador se entrevistó con el director y verificó si había sido informado previamente de la visita. Con su apoyo reunió a los maestros de segundo, quinto y sexto grados, les entregó el cuestionario destinado a ellos, acordó la hora y lugar de la entrevista al maestro seleccionado y acordó con el o los docentes de sexto grado la aplicación del cuestionario a sus alumnos. Al cabo de estas dos últimas actividades recogió los instrumentos entregados a los docentes.

En el mes de mayo se capturó la información recogida con los cuestionarios y entre junio y julio se transcribieron las entrevistas realizadas.

Referencias

- Black, Paul, y D. William (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, Vol. 5 (1): 7-75.
- Brookhart, Susan M. (2007). Expanding Views About Formative Assessment: A Review of the Literature. En McMillan, James H. Ed. *Formative Classroom Assessment: Theory into Practice*. New York. Teachers College Press, pp. 43-62.
- CERI (2005). *Formative assessment. Improving learning in secondary classrooms*. Paris. OECD.
- Díaz Barriga, Ángel *et al.* (2010). *Módulo 3: Evaluación para el aprendizaje en el aula*. Reforma Integral de la Educación Básica 2009. Diplomado para Docentes de Primaria. México, SEP.
- Dunn, Karee E. y Sean W. Mulvenon (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment Research and Evaluation*. Vol. 14 (7).
- Resnick, L. B. *et al.* (2010). The Relationship between Assessment and the Organization and Practice of Teaching. En Peterson, Penelope, E. Baker y B. McGaw, eds. (2010). *International Encyclopedia of Education*, 3rd Ed. Amsterdam, Elsevier-Academic Press, Vol. 3, pp. 397-402.
- Ruiz-Primo María Araceli *et al.* (2010). Formative Assessment, Motivation and Science Learning. En Andrade, Heidi L. y Gregory J. Cizek, Eds. (2010). *Handbook of formative assessment*. New York-London, Routledge, pp. 139-158.
- Schneider, M. Christina y Bruce Randel (2010). Research Characteristics of Effective Professional Development Programs for Enhancing Educators' Skills in FA. En Andrade y Cizek, 2010: 251-276.
- Stiggins, Richard J., J. Arter, (2002). *Assessment for Learning, International Perspectives*. International Conference on Assessment for Learning. Portland. Assessment Training Institute.

Agradecimientos

Se agradece el apoyo del CONACYT al proyecto (Convocatoria SEB 2008, 000110092).