

## ESTUDIOS SOBRE PRÁCTICAS DE EVALUACIÓN EN AULA: REVISIÓN DE LITERATURA

---

FELIPE MARTÍNEZ RIZO / ADRIANA MERCADO SALAS  
Universidad Autónoma de Aguascalientes

**RESUMEN:** Algunos esfuerzos por mejorar la calidad educativa parten de la idea de que el rendimiento de los alumnos mejora si el maestro utiliza evaluaciones de enfoque formativo, pero las investigaciones al respecto no son concluyentes. Sin embargo, en esos estudios muchas veces la práctica de evaluación ha sido caracterizada como formativa o no formativa verificando sólo la presencia o ausencia de ciertas conductas, sin considerar aspectos más finos, como el nivel de demanda cognitiva con que se

manejan los contenidos o la forma de dar retroalimentación. Este trabajo sintetiza los resultados de algunos de los pocos trabajos que han buscado expresamente estudiar en detalle las prácticas de evaluación en aula.

**PALABRAS CLAVE:** Práctica docente, evaluación en aula, evaluación formativa, educación básica, estado del conocimiento.

### Introducción

La revisión de estudios empíricos sobre el impacto de las prácticas de Evaluación Formativa (EF) en el aprendizaje (Martínez Rizo 2012) encontró que algunos de los trabajos que concluyen que dicho impacto es muy grande tienen limitaciones metodológicas que ponen en duda sus conclusiones (Black y Wiliam, 1998; Dunn y Mulvenon, 2009; Kingston y Nash, 2011a y 2011b; Briggs et al. 2012).

Las críticas centran la atención en las limitaciones de las técnicas de meta-análisis utilizadas para integrar los resultados de trabajos previos, pero estos tienen otra falla que puede explicar mejor la escasa consistencia de los resultados sobre el impacto de la EF: el grado en que las prácticas de evaluación en aula estudiadas tenían o no en realidad los rasgos propios de una EF. Algunos estudios sobre la EF presentan poco rigor para estudiar las prácticas, al basarse simplemente en la declaración de los maestros mismos.

Esta revisión analiza algunos de los pocos estudios centrados en el análisis de las prácticas de evaluación en aula, buscando identificar sus alcances y límites metodológicos.

## Los trabajos de Stiggins y colaboradores

La obra presenta trabajos dirigidos por Stiggins y Conklin (1992), describiendo cómo evolucionaron los acercamientos metodológicos para estudiar las prácticas de evaluación a medida que se quería estudiar aspectos más complejos:

- Encuesta a 228 maestros y análisis de diarios escritos por 32 de ellos.
- Observación participante de tres aulas de primaria durante 10 semanas.
- Dos estudios en ocho salones para sistematizar las dimensiones de las prácticas de evaluación en aula y hacer perfiles detallados de ellas.
- Estudio con 36 maestros sobre evaluaciones de alto nivel cognitivo.
- Estudio de las prácticas de calificación de 15 maestros, contrastando las recomendaciones al respecto de los libros de texto con la práctica real.

Según la encuesta con 228 maestros, en primaria la proporción que dijo elaborar sus pruebas era de 34% en primaria y bajaba a 11% en secundaria. Alrededor de la mitad se sentían cómodos con esos instrumentos y de 9 a 14% consideraba dominarlos; 42% dijeron estar preocupados por mejorar sus propias pruebas y 19% dijo no tener tiempo para desarrollarlas (Stiggins y Conklin, 1992: 35-41).

El análisis de diarios de profesores permitió analizar 290 actividades de evaluación y mostró que en 65% de los casos el propósito era calificar, en 18% diagnosticar y en 6% retroalimentar. Sólo en tres casos se evaluaban habilidades de alto nivel cognitivo. En 40% de los casos las evaluaciones fueron pruebas; el 60% restante consistía en observaciones de los alumnos o de sus trabajos. Hubo consenso en que las pruebas externas eran limitadas y poco útiles para dar información diagnóstica relevante alineada con el currículo (Stiggins y Conklin, 1992: 42-52).

Las observaciones evidenciaron la complejidad de las prácticas de evaluación. A juicio de un maestro, los exámenes externos eran rutinarios y evaluaban memoria, en tanto que los que él mismo hacía tenían preguntas de todos los niveles de demanda

cognitiva, pero el análisis de esos exámenes mostró que casi todas sus preguntas se enfocaban sólo a memorización (Stiggins y Conklin, 1992: 105-115).

Los trabajos en clase también incluían principalmente cuestiones de memorización y no había actividades que implicaran análisis o inferencia. Esta tendencia era más marcada en preguntas orales: el maestro exponía un tema con datos puntuales y luego preguntaba los mismos datos (Stiggins y Conklin, 1992: 109 y 117).

Las pruebas estandarizadas eran consideradas confiables y los maestros trataban de imitarlas, mientras la observación del desempeño y las preguntas tipo ensayo eran vistas como subjetivas y menos aceptables, por lo que su uso era mínimo. En cuanto a retroalimentación, la mayoría de las tareas eran marcadas con símbolos y se daban a los alumnos oportunamente (Stiggins y Conklin, 1992: 142-143).

En la revisión de trabajos previos sobre la evaluación de habilidades de alto nivel cognitivo, se encontró que Fleming y Chambers (1983) reportan que, de cerca de 9000 preguntas elaboradas por los maestros, más del 90% se enfocaba solo al recuerdo de hechos puntuales (Stiggins y Conklin, 1992: 154-155).

El trabajo con 36 maestros mostro que la parte de preguntas de memorización fue de 56% en primaria y 41% en secundaria, y que las que implicaban inferencia fueron 19% en el primer caso y 44% en el último. De las preguntas, en primaria el 70% correspondió a memorización y el 17% a inferencia, en tanto que en secundaria la proporción fue 42% y 18% (Stiggins y Conklin, 1992: 161-163).

Contra lo que recomiendan los textos, los 15 maestros del último estudio dijeron que el esfuerzo se debe tener en cuenta al calificar, además del desempeño efectivo, y la mitad usó formas de calificar distintas para alumnos que consideraba más o menos capaces. Nueve de 15 califican todos los trabajos de los alumnos, pese a que los textos recomiendan que muchos tengan un papel sólo formativo. Prácticamente no hubo evidencias de un cuidado sistemático de los profesores al calificar; la forma de combinar resultados parciales y llegar a una puntuación numérica también fue muy arbitraria (Stiggins y Conklin, 1992: 170-171).

## El proyecto en países de América Latina

El estudio incluyó ocho países, 80 escuelas y 160 maestros de 6° grado, a los que se aplicó un cuestionario y una entrevista. Se tomaron 4360 registros fotográficos de evaluaciones y trabajos de alumnos y se analizaron los currículos nacionales.

El discurso docente se centra en las funciones básicas de la evaluación (sumativa y formativa), de manera poco específica, limitándose a caracterizar la evaluación de modo general; en algunos casos hubo contradicciones en la entrevista, lo que revela poca solidez de los conceptos expresados (Picaroni, 2009: 41).

El instrumento más utilizado es la prueba objetiva, en forma mensual, bimensual o trimestral, principalmente para calificar. Se usan instrumentos de aplicación diaria: tareas, ejercicios, hojas de trabajo y actividades orales (Picaroni, 2009: 29-30).

Se valora la disposición y compromiso de los alumnos, la puntualidad, el interés, la disciplina y la limpieza. La mayoría de los docentes no explicita el peso de cada elemento, pero enfatiza que el examen tiene más (Picaroni, 2009: 35).

Las decisiones de calificación, aprobación o reprobación no se basan en criterios homogéneos, pese a que están plasmados en los objetivos del programa de cada materia, por lo que al fin de año no es posible garantizar que todos los alumnos sepan lo mismo, aunque tengan el mismo promedio (Picaroni, 2009: 56).

Las consignas empleadas en las propuestas de evaluación suelen estar cargadas de ambigüedades y aspectos implícitos. La mayoría son puramente escolares, descontextualizadas y sin audiencias plausibles; son excepcionales las propuestas de evaluación basadas en actividades auténticas (Picaroni, 2009: 62 y 42).

En cuanto a retroalimentación a los alumnos, pocos docentes hacen devoluciones en sentido estricto. Aunque algunos docentes manejan la distinción entre evaluación formativa y sumativa, y dicen preferir la primera, en la práctica muchos califican todos los trabajos y no dan otro tipo de devolución a los alumnos más allá de la calificación. En el mejor de los casos reemplazan la calificación por juicios del tipo “muy bien” o “debes esforzarte más” (Picaroni, 2009: 85 y 100).

Las devoluciones que los maestros hacen a los niños a partir de sus trabajos no los orientan sobre las formas para avanzar. Se limitan a asignar calificaciones con números o letras sin informar sobre los aspectos involucrados ni indicar formas de mejorar. Muchos se preocupan por estimular a los alumnos para que se esfuercen, pero no les dan pistas sobre cómo hacerlo (Ravela, Picaroni y Loureiro, 2009: 5).

Para enfrentar la situación de los alumnos de resultados más bajos los docentes apelan en general a factores emocionales y al esfuerzo personal de esos alumnos, sin hacer ajustes en su propia práctica (Picaroni, 2009: 101-102).

Los padres de familia deberían recibir información que permita apoyar mejor a sus hijos, pero casi siempre la información que reciben las familias sobre el logro de sus hijos está centrada en la calificación, que sólo dan cuenta de una posición relativa en una escala de valoración (Ravela, Picaroni y Loureiro, 2009: 5).

En algunos países estudiados las evaluaciones nacionales están más avanzadas y en línea con los desarrollos de la didáctica que las propuestas de los docentes. Este hallazgo contradice la extendida idea de que las evaluaciones externas son solo memorísticas. Si esto es verdad, las evaluaciones externas podrían apoyar la tarea de los docentes, si las aprovechan como herramientas al servicio de su propia propuesta educativa (Ravela, Picaroni y Loureiro, 2009: 11-12).

## Estudios mexicanos

### ***El trabajo de Ismael Vidales Delgado y colaboradores***

El texto recoge resultados de un proyecto realizado en el estado de Nuevo León. Un análisis documental fue seguido por entrevistas con informantes clave, y luego se trabajó con una muestra de 313 docentes de primarias públicas.

Según este trabajo, la idea que tienen los maestros sobre la evaluación es fragmentada y dispersa, limitándose a reproducir la terminología de los textos que se manejan en cursos de actualización. Para elaborar instrumentos o calificar, los maestros tienden a guiarse por su experiencia más que por principios teóricos, que suelen estar ausentes, en opinión de supervisores y jefes de sector. Algunos piensan que no hay tal

desconocimiento teórico, sino más bien resistencias debido al exceso de carga administrativa (Vidales Delgado et al., 2005: 184).

El discurso muchos maestros coinciden en que *se evalúa para mejorar la práctica docente y los desempeños de los alumnos*. En la práctica prevalece la visión de que la evaluación es un evento ocasional que sirve para sustentar juicios sobre aprobación o reprobación de los alumnos, *un requisito para dar resultados numéricos a la administración*. Estas opiniones son reforzadas por una cultura en la que padres y autoridades ven la evaluación como formalismo que da evidencias cuantitativas del avance de sus hijos (Vidales Delgado et al., 2005: 185).

Según la opinión de supervisores y jefes de sector, un alto porcentaje de escuelas y maestros (hasta 80%) compran exámenes a editoriales y otras instancias, pero 72% de los directores sostiene que en sus escuelas los exámenes son hechos por los maestros, y sólo 28% reconoce que los adquieren de instancias externas.

En cuanto a la información para los padres, no hay informes amplios ni se da seguimiento a los resultados del proceso; *sólo se dan las boletas de calificación y las observaciones de felicitación o descalificación* (Vidales Delgado, 2005: 194).

### ***El estudio del Instituto Nacional para la Evaluación de la Educación***

Entre 2009 y 2010 se aplicó un cuestionario sobre prácticas de evaluación a 3534 profesores de 1193 escuelas, otro a 17908 alumnos y otro a los directores de los mismos planteles. La muestra de escuelas fue representativa a nivel nacional.

La mayoría de docentes de primaria dicen utilizar la evaluación para mejorar el aprendizaje y la enseñanza, en forma congruente con un enfoque formativo. Un 68% aplica exámenes bimestrales; 31% lo hace en forma mensual, quincenal o semanal y menos de 1% no los utiliza (García Medina et al., 2011: 28 y 71-72).

Unos datos van en el sentido de que los maestros usarían con más frecuencia actividades de evaluación de complejidad alta y media, y con menor frecuencia memorización (García Medina et al., 2011: 52-56), lo que no es consistente con datos sobre la calificación, según los cuales los aspectos más importantes para dar una nota

fueron atención y participación en clase (71%), y calificación en los trabajos en clase (48%). Además 63% de los maestros califica “considerando aciertos y errores” de forma cotidiana (García Medina *et al*, 2011: 55 y 77).

Tampoco parece consistente que 70% de los profesores utilicen con frecuencia firmas o sellos como medio de retroalimentación, y que una proporción similar diga que explica diariamente cómo solucionar los errores cometidos al retroalimentar a sus estudiantes (García Medina *et al*, 2011: 76-79).

### ***Diagnóstico de prácticas de evaluación de maestros de Nuevo León***

El estudio se hizo con una muestra de primarias públicas, con 96 escuelas, 409 docentes y 3,358 alumnos (Ruiz Cuellar y Perez Martinez, en prensa).

Una proporción alta de maestros informa que diario o casi diario revisa tareas y observar el trabajo de los alumnos. Más del 80% dice hacer preguntas orales o pedir trabajos desde dos veces a la semana hasta quincenalmente; 82% de los alumnos dice tener exámenes bimestrales, 26% cada mes y 21% cada semana.

En la mayor parte de las escuelas los exámenes son diseñados por los docentes. Siete de cada diez maestros revisa y califica los exámenes que aplica. Para llegar a la calificación bimestral, el aspecto más importante es el desempeño en clase, seguido por los ejercicios en el cuaderno, el portafolio y los ejercicios de los libros de texto. Nueve de cada diez docentes dice informar a sus alumnos sobre los aprendizajes a evaluar y los criterios de evaluación.

La compra de exámenes, según los docentes, ocurre sólo en una tercera parte de los casos y la relación de las preguntas del examen con los temas vistos en clase es muy alta; 50% opina que casi todas las preguntas muestran correspondencia y 35% que sucede con la mayor parte de preguntas. Un 78% de los alumnos dice que las preguntas en su mayoría son fáciles, pero hay algunas difíciles y 13% que prácticamente todas las preguntas son fáciles. En cuanto a las pruebas externas, la mayoría de los alumnos dicen que son preparados para ellas por sus maestros.

## Conclusión

Los trabajos mexicanos coinciden con los latinoamericanos y los estadounidenses revisados. En todos, la visión que tienen de su práctica los maestros concuerda con los principios de evaluación formativa, pero hay elementos que hacen temer que las prácticas no sean congruentes con las concepciones y las creencias:

- Que los obstáculos para implementar prácticas formativas según los maestros sean principalmente el tamaño de los grupos y la falta de tiempo.
- La excesiva importancia que siguen dando los maestros a la calificación de todos o casi todos los instrumentos de evaluación que utilizan.
- Cierta contradicción en las percepciones de los maestros sobre sus propias competencias para hacer evaluación formativa: predomina una visión muy positiva sobre las competencias que se tienen, pero la mayoría coincide en que es necesario o muy necesario recibir apoyo o formación para desarrollar mejor esta dimensión del quehacer docente.

Por otra parte, parecería haber una buena alineación entre enseñanza y evaluación, pero hay también bases para plantear una hipótesis en cuanto a la predominancia de propósitos educativos de baja demanda cognitiva como posible explicación. Un indicio es la masiva opinión de los alumnos de que los exámenes que responden son fáciles, que puede indicar que las preguntas tienen bajo nivel de demanda cognitiva. La alineación entre enseñanza y evaluación puede ser real, con tareas de baja demanda cognitiva en una y otra. De ser así, muchos hallazgos cobrarían sentido. Corroborar esta hipótesis será tarea de futuros trabajos.



## Referencias

- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*. Vol. 5 (1): 7-75.
- Briggs, D., Ruiz-Primo, A., Furtak, E., Shepard, L. & Yin, Y. (2012). Meta-Analytic Methodology and Inferences about the Efficacy of Formative Assessment. *Educational Measurement: Issues and Practice*. Vol. 31 (4): 13-17.
- Dunn, K. & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. *Practical Assessment Research & Evaluation*, 14(7). Available online: <http://pareonline.net/getvn.asp?v=14&n=7>
- Fleming, M. & Chambers, B. (1983). Teacher-made tests: Window on the classroom. En W. E. Hathaway, ed., *Testing in the schools. New directions for testing and measurement*. San Francisco, Jossey Bass, 19: 29-38.
- García Medina, A. M. et al. (2011). *Evaluación de los aprendizajes en el aula. Opiniones y prácticas de docentes de primaria en México*. México: Instituto Nacional para la Evaluación de la Educación.
- Kingston, N. & Nash, B. (2012). How Many Formative Assessment Angels can Dance on the Head of a Meta-Analytic Pin. *Educational Measurement: Issues and Practice*. Vol. 31 (4): 18-19.
- Kingston, Neal y B. Nash (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*. Vol. 30 (4): 28-37.
- Loureiro Denis, G. (2009). *Evaluación en el aula, currículo y evaluaciones externas*. Instituto de Evaluación Educativa, UCU y GTEE-PREAL.
- Martínez Rizo, F. (2012). Investigación empírica sobre el impacto de la EF. *Revista Electrónica de Investigación Educativa*. Vol. 14 (1), 1-15. ISSN 1607-4041.
- Picaroni, B. (2009). *La evaluación en las aulas de primaria: usos formativos, calificaciones y comunicación con los padres*. Instituto de Evaluación Educativa, UCU y GTEE-PREAL.
- Ravela, P., Picaroni, B. & Loureiro, G. (2009). La evaluación de aprendizajes en las aulas de 6° grado en América Latina. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 12 (Julio), pp. 1-12. GTEE-PREAL.
- Ruiz Cuéllar, G. & Perez Martínez, G. (en prensa). Creencias y prácticas de evaluación en aula de maestros de educación primaria de Nuevo León. En Ruiz Cuéllar, Guadalupe, ed. *La evaluación en el aula: Diagnóstico*. Cap. 2. Aguascalientes. Universidad Autónoma de Aguascalientes.

Stiggins, R. & Conklin, N. (1992). *In teacher's hands: Investigating practices of classroom assessment*. Albany, State University of New York Press.

Vidales D., Isamael et al. (2005). *Prácticas de evaluación escolar en el nivel de educación primaria en el Estado de Nuevo León*. CAEIP. México: Santillana.