

Revista Electrónica de Investigación Educativa

Vol. 17, Núm. 1, 2015

Estudios sobre prácticas de evaluación en el aula: revisión de la literatura

Research on Classroom Assessment Practices: A Literature Review

Felipe Martínez-Rizo (*)
fmrizo@prodigy.net.mx

Adriana Mercado (*)
amercado@correo.uaa.mx

(*) Universidad Autónoma de Aguascalientes

(Recibido: 30 de mayo de 2013; aceptado para su publicación: 8 de octubre de 2014)

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 considerando 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 de los maestros. Los resultados orientarán el trabajo de diseño de instrumentos que permitan obtener información de mejor calidad que la habitual sobre las prácticas, como busca la línea de investigación de la que forma parte este trabajo.

Palabras clave: Práctica educativa, evaluación formativa, educación básica, estado del conocimiento.

Abstract

Some efforts to improve the quality of education are based on the idea that student performance improves if teacher uses formative assessment, but research is inconclusive about it. However, in many projects assessment practice has been characterized as formative or not considering only the presence or absence of previously specified behaviors, without considering finer aspects, such as the level of cognitive demand of content, or the way teachers give feedback. This paper summarizes the results of projects that explore in detail teachers' classroom assessment practices. The results will guide the design of tools to gather good quality information about teachers' practices.

Keywords: Educational practice; formative evaluation, elementary education; literature review.

I. Introducción

En la revisión de estudios empíricos sobre el impacto de las prácticas de Evaluación Formativa (EF) en el aprendizaje de los alumnos (Martínez Rizo, 2012a y 2012b) se ha encontrado que algunos de los trabajos más conocidos, que concluyen que dicho impacto es muy grande incluso en períodos cortos de tiempo, adolecen de limitaciones metodológicas claras, que ponen en duda sus conclusiones. Esto no implica que el impacto real sea nulo o negativo, pero sí que se necesitan trabajos más rigurosos para resolver un debate aún abierto (Black y William, 1998; Dunn y Mulvenon, 2009; Kingston y Nash, 2011; Briggs, Ruiz-Primo, Furtak, Shepard y Yin, 2012).

La polémica ha centrado la atención en las limitaciones de las técnicas de meta-análisis utilizadas para integrar los resultados de trabajos previos sobre el tema, pero estos trabajos presentan otra falla a la que se presta poca atención y podría explicar mejor la escasa consistencia de los estudios sobre el impacto de la EF: el grado en que las prácticas de evaluación en aula que se manejaron en esos estudios tenían o no en realidad los rasgos propios de una EF.

Un buen diseño de investigación para explorar la posible influencia de la EF sobre el aprendizaje no sólo debe controlar la influencia de otras variables; debe también medir de la manera más precisa posible tanto la variable dependiente –el aprendizaje– como la independiente, el grado en que las prácticas de evaluación de los docentes tienen realmente los rasgos que permitan catalogarlas como EF. En muchos casos, sin embargo, las prácticas de los docentes no se miden con un mínimo de rigor, sino que se les clasifica como EF con base en la declaración de los mismos maestros, aunque no necesariamente refleja su práctica real.

Los estudios empíricos enfocados a explorar las prácticas de evaluación de los docentes son escasos, tal vez porque la complejidad de la práctica docente hace que en muchos casos no se incluya entre los aspectos explorados. Esto llevó a realizar una revisión de literatura de textos con resultados de un subconjunto de trabajos empíricos

sobre las prácticas de evaluación en aula de los maestros:

- Un libro de Stiggins y Conklin (1992) que recoge trabajos realizados en el sistema educativo de Estados Unidos entre 1980 y 1990.
- Tres textos derivados de un estudio de 2008 sobre las prácticas de evaluación en las primarias de ocho países de América Latina (Loureiro, 2009; Picaroni, 2009; Ravela, Picaroni y Loureiro 2009).
- Tres textos de investigaciones de México: Vidales y Elizondo (2005); García Medina et al. (2011), y Ruiz y Pérez (en prensa).

Aunque en los últimos años está aumentando el número de trabajos sobre las prácticas docentes y, en particular, las prácticas de evaluación en aula, el tema es relativamente reciente y se ha trabajado sobre todo en países anglosajones. Los trabajos al respecto realizados en países de habla hispana son particularmente escasos, aunque la literatura sobre evaluación en general sea abundante.

Este trabajo pretende contribuir al desarrollo del campo al revisar trabajos relativamente antiguos, pero que ilustran la forma en que se desarrolló el tema en el medio estadounidense hace un par de décadas (Stiggins, Arter, Chappuis y Chappuis, 2007), así como recientemente en América Latina y México.

Los trabajos de Stiggins y Conklin

En *In teachers' hands*, Stiggins y Conklin (1992) describen cómo evolucionaron los acercamientos para estudiar las prácticas de evaluación, a medida que se advertía la necesidad de captar de manera confiable aspectos cada vez más complejos:

- Primero se realizó una revisión de los pocos antecedentes que se pudieron identificar en la literatura. (Cap. 2, pp. 9-30).
- Luego se hizo una encuesta con 228 maestros de ocho distritos de distintas regiones de Estados Unidos y se pidió a 32 de ellos que llevaran un diario durante diez semanas; recogiendo la versión que los docentes dieron de su propia práctica, con información para indagar si corresponde con lo que realmente hacen. (Cap. 3, pp. 31-52).
- Una observación participante del trabajo docente en tres aulas de sexto de primaria, durante 10 semanas del último trimestre del año, de manera que se asegurara observar cada salón durante al menos 20 días y se incluyeran todos los días de la semana. (Cap. 4, pp. 53-77).
- Dos estudios más en ocho aulas de educación media para construir un marco analítico que sistematizara las dimensiones del constructo "prácticas de evaluación en aula". (Cap. 5, pp. 79-98).
- Con ese marco analítico se hicieron perfiles detallados de las actividades de

evaluación en las ocho aulas. El Capítulo 6 contrasta una clase de historia y otra de economía, mientras en el Capítulo 7 se sintetizan las semejanzas y diferencias de los ocho perfiles estudiados.

- Otro estudio exploró el grado en que los maestros evaluaban diariamente habilidades de alto nivel cognitivo. Se trabajó con 36 maestros voluntarios, recogiendo de cuatro a seis ejemplos de evaluaciones recientes aplicadas por los maestros y se les entrevistó en relación con ellas, además de observarlos durante un día. (Cap.8, pp. 154-167)
- Un trabajo más exploró las prácticas de calificación de 15 maestros, a partir del contenido de algunos textos sobre evaluación que se utilizan en muchos programas de formación, para contrastar sus recomendaciones con la práctica real. (Cap.8, pp. 167-175)

La revisión del capítulo 2 confirmó que el ambiente de la década de 1980 estaba dominado por las pruebas estandarizadas en gran escala, al grado que:

(...) uno puede especular que las prioridades de la evaluación en gran escala han dominado el campo tan completamente y durante tanto tiempo, que los académicos no han logrado siquiera percibir o reconocer que puede haber otras prioridades relacionadas con tipos diferentes de evaluación, como la que tiene lugar en las aulas. (Stiggins y Conklin, 1992, p. 9).

Los autores señalan que, pese a su propia capacidad para hacer buenas evaluaciones en gran escala, no sabían cómo apoyar a los maestros *a enfrentar las demandas de la medición del rendimiento de sus alumnos día a día*, aunque esto es una dimensión fundamental de la enseñanza. Después de revisar informes de estudios sobre pruebas en gran escala y sobre enseñanza, concluyeron que prácticamente no se prestaba atención a las prácticas de evaluación del docente.

Para evaluar a sus alumnos los docentes no dan gran importancia a las pruebas, sino a sus propias observaciones; en primaria una tercera parte elabora pruebas estandarizadas; en educación media, alrededor de 75%. Los maestros dicen tener dificultad para elaborar ese tipo de instrumentos. (Stiggins y Conklin, 1992, pp. 12-20)

Según la encuesta respondida por 228 maestros, la proporción que informa no usar pruebas elaboradas por ellos mismos era de 34% en los primeros grados de primaria y bajaba a 11% al fin de la enseñanza media, en tanto que los que dicen no usar pruebas ajenas, que al principio de primaria era también 34%, aumentan hasta 52% al fin de la enseñanza media. En todos los casos alrededor de la mitad (49-52%) se sentía cómodo al usar esos instrumentos y de 9 a 14% consideraba dominarlos. Sin embargo 42% dijo estar preocupado por mejorar sus propias pruebas y otro 19% dijo no tener tiempo para desarrollarlas. Un 28% comentó no tener preocupaciones al respecto. (Stiggins y Conklin, 1992, pp. 35-41).

El análisis de diarios de 32 profesores comprendió 290 actividades de evaluación y mostró que, en dos terceras partes, el propósito era calificar o determinar si los alumnos dominaban ciertos contenidos, en 18% diagnosticar y retroalimentar sólo en

6%. Sólo en tres casos se evaluaban habilidades de alta complejidad cognitiva. 40% de las evaluaciones fueron pruebas (24% hechas por el docente); el 60% restante consistía en observar el comportamiento de los alumnos o sus trabajos. En general los maestros se mostraban dispuestos a aceptar que sus evaluaciones tenían puntos débiles, pero no podían analizar en detalle esas debilidades. Se manifestó consenso de que las pruebas eran limitadas y poco útiles para dar información diagnóstica alineada al currículo. (Stiggins y Conklin, 1992, pp. 42-52)

Los estudios cualitativos pusieron en evidencia la complejidad de las prácticas de evaluación, que los trabajos anteriores no conseguían abarcar. Esos trabajos se hicieron en escuelas de educación media y sus resultados son consistentes con los de trabajos anteriores en los aspectos más gruesos, pero aportaron elementos nuevos, en particular sobre los niveles de demanda cognitiva de las evaluaciones.

En el capítulo 6 se describe el trabajo de un maestro de historia que consideraba que los exámenes de los textos eran rutinarios y evaluaban memoria, en tanto que los que elaboraba él mismo 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, pp. 105-106 y 115).

Los trabajos que ponía en clase también incluían principalmente cuestiones de memorización (70%), 20% implicaban hacer comparaciones y 10% evaluar textos; estaban ausentes actividades que implicaran análisis o inferencia. Esta tendencia era más marcada en preguntas orales. El maestro solía exponer un tema con datos puntuales y después preguntaba los mismos datos en los exámenes. Las tareas también se referían a hechos específicos, sobre todo las que se hacían para preparar a los alumnos para un examen. (Stiggins y Conklin, 1992, pp. 109 y 117).

Luego se compara el profesor descrito con otros en varios ambientes de trabajo (asignaturas y grados). Los exámenes y las tareas de otros maestros en general tenían buen nivel, salvo la evaluación de habilidades de alto nivel cognitivo, que en la mayoría de los casos tenía criterios de desempeño y procedimientos de calificación definidos de manera vaga. (Stiggins y Conklin, 1992, pp. 141).

Las pruebas estandarizadas tendían a ser consideradas confiables y los maestros trataban de imitarlas, en tanto que las evaluaciones basadas en observación del desempeño y en preguntas de tipo ensayo eran vistas como subjetivas y menos aceptables, y su uso era mínimo. Las actividades de evaluación ocupaban buena parte del tiempo. El docente promedio decía dedicar alrededor de una cuarta parte del tiempo de clase a ellas y el resto a exposición de temas (40%), instrucción individual (25%) y planeación (10%). (Stiggins y Conklin, 1992, p. 142 y 145).

Por lo que se refiere a la retroalimentación a los alumnos sobre los resultados de las evaluaciones, las tareas en su mayoría eran marcadas con un símbolo y los alumnos recibían esa información oportunamente. (Stiggins y Conklin, 1992, p. 143).

La mayoría de los maestros consideraba que enseñanza y evaluación eran cosas

separadas, sin relación entre sí y no sabían cómo hacer buen uso de la evaluación durante la enseñanza para hacerla más efectiva. Los maestros dijeron pasar poco tiempo reflexionando sobre la naturaleza o calidad (validez, confiabilidad y valor comunicativo) de sus evaluaciones (Stiggins y Conklin, 1992, pp. 148 y 141).

En cuanto a la evaluación de habilidades de alto nivel cognitivo, la revisión de trabajos previos muestra que los maestros encuentran difícil redactar preguntas de alta demanda cognitiva (Carter, 1984). Según Fleming y Chambers (1983), de cerca de 9,000 preguntas elaboradas por los maestros, más de 90% se enfocaba al recuerdo de hechos puntuales. Según Reynolds y Menard (1989) los maestros ignoran los niveles taxonómicos de Bloom. La proporción de preguntas de memorización fue de 56% en los primeros grados de primaria y bajó a 41% en los últimos de enseñanza media; las que implicaban inferencias subieron de 19% en el primer caso a 44% en el último. De las preguntas orales, en los primeros grados de primaria el 70% fue de memorización y el 17% de inferencia, en tanto que en los últimos grados de enseñanza media las proporciones respectivas fueron 42% y 18%. (Stiggins y Conklin, 1992, p. 154-155 y 161-163).

Los 15 maestros del último estudio consideraron que el esfuerzo del alumno se debe tener en cuenta al calificar su trabajo, además de su desempeño. La mitad manejó formas de calificar distintas para alumnos más o menos capaces. 9 de los 15 profesores califican todos los trabajos de los alumnos y no se cuida bien la calidad de la calificación; la forma de combinar resultados parciales y llegar a una puntuación numérica fue muy arbitraria. (Stiggins y Conklin, 1992, pp. 170-171).

II. El proyecto latinoamericano

El estudio fue realizado en ocho países e incluyó 80 escuelas y 160 maestros de 6o. grado, a los que se aplicó un cuestionario y una entrevista semiestructurada. Se tomaron 4,360 registros fotográficos de propuestas de evaluación y trabajos de estudiantes y se hizo un análisis de los currículos nacionales de los ocho países que participaron. Los resultados del estudio se han difundido en tres textos:

- Boletín Núm. 12, *Observatorio Regional de Políticas de Evaluación Educativa* (Ravela, Picaroni y Loureiro, 2009), que presenta sintéticamente el estudio, sus objetivos y metodología y los resultados más importantes.
- Picaroni (2009) que muestra hallazgos en relación con el enfoque y discurso del maestro sobre la evaluación y su práctica, la retroalimentación que da a los alumnos y la comunicación de resultados a sus padres.
- Loureiro (2009) que hace un análisis de las interrelaciones o congruencias que hay entre los currículos nacionales, lo que evalúan los maestros y lo que evalúan en los países con las pruebas externas estandarizadas.

En cuanto al tipo de evaluación que realizan, muchos docentes manifiestan hacer una diagnóstica al inicio del año lectivo, para tomar decisiones de planificación, así como

una evaluación al final de cada bimestre. (Picaroni, 2009, p. 19).

El discurso se centra en las funciones básicas de la evaluación (sumativa y formativa), pero de manera poco específica. Los docentes se limitan a caracterizar las evaluaciones de modo general, y en algunos casos incurren en contradicciones a lo largo de la entrevista, situación que permite suponer poca solidez de los conceptos expresados. (Picaroni, 2009, p. 41).

El instrumento más mencionado es la prueba objetiva o examen; es usado todos los meses, o en forma bimensual o trimestral, principalmente para calificar a los alumnos. Se utilizan instrumentos de aplicación diaria (tareas, ejercicios, hojas de trabajo, actividades escritas u orales) que conforman lo que se denomina "evaluación continua". En la mayoría de los casos, se realizan en el cuaderno de clase o en forma oral, con apoyo del pizarrón. (Picaroni, 2009, pp. 29-30).

En el contexto educativo mexicano se valora la disposición y el compromiso de los alumnos, la puntualidad, el interés en las actividades realizadas, la disciplina y la limpieza de los trabajos. La mayoría de los docentes no explicita la ponderación de cada uno de estos elementos, pero enfatiza que el examen es lo que tiene más peso a la hora de promediar los puntos obtenidos por el trabajo realizado a lo largo de un bimestre. (Picaroni, 2009, p. 35).

La toma de decisiones de calificación, aprobación o reprobación no se fundamenta en criterios homogéneos entre los docentes, pese a que existen y están plasmados en los objetivos de los programas escolares para cada materia y grado. En México los profesores tienen libertad para decidir mediante qué actividades los alumnos demostrarán su dominio de los contenidos vistos en cada materia y su peso, por lo que al fin de año no es posible garantizar que todos tengan el mismo nivel de conocimientos, aunque tengan el mismo promedio. (Picaroni, 2009, p. 56).

Las normas escolares suelen establecer en forma general objetivos programáticos pero de manera poco precisa. En ninguno de los países estudiados las normas incluyen un respaldo conceptual de carácter descriptivo que asegure la relación entre la calificación otorgada a un alumno y los desempeños a los que ella refiere. Ni en las normas de evaluación ni en los currículos se explicitan en detalle los referentes de la evaluación. (Picaroni, 2009, p. 60).

Las consignas empleadas en las propuestas de evaluación suelen estar cargadas de ambigüedades y aspectos implícitos, que tal vez sean claros para el docente pero seguramente no lo son para el alumno. Por otra parte, la mayoría de las propuestas son puramente escolares, descontextualizadas y sin audiencias medianamente plausibles; son excepcionales las propuestas de evaluación basadas en actividades auténticas. (Picaroni, 2009, pp. 62 y 42).

En cuanto a retroalimentación a los alumnos, pocos docentes hacen devoluciones en sentido estricto, es decir expresiones que permitan al alumno comparar lo efectivamente realizado con lo que se esperaba que lograra, a fin de comprender las

insuficiencias de su trabajo y profundizar en el conocimiento cuando los trabajos están bien logrados. 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, pp. 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, p. 5).

Además de orientar los esfuerzos de los estudiantes, la evaluación debe servir para que los maestros enfoquen mejor su propio trabajo, a partir del avance de sus alumnos, considerando en particular a los que tienen dificultad para aprender. Pocos docentes desarrollan estrategias didácticas a partir de las necesidades cognitivas detectadas gracias a los resultados de los alumnos; en general para enfrentar la situación de los alumnos de resultados más bajos los docentes apelan a factores emocionales y al esfuerzo personal de esos alumnos, sin hacer ajustes en su propia práctica. (Picaroni, 2009, pp. 101-102).

Otros destinatarios de la retroalimentación son los padres de familia, que deberían recibir información que permita apoyar mejor a sus hijos. Sin embargo el trabajo mostró que hay pocas diferencias en los países estudiados en este sentido: las diferencias se refieren a aspectos menos sustantivos, como la forma de expresar la calificación. 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, p. 5).

Los currículos de los países estudiados se enfocan al desarrollo de competencias y habilidades; destacan la importancia de no limitar la educación a la transmisión de conocimientos que deberían aplicarse en situaciones varias en la vida diaria; describen a la evaluación como un proceso que debe ser continuo, permanente y flexible; y señalan la necesidad de elaborar evaluaciones diagnósticas, además de formativas y sumativas. (Loureiro, 2009, p. 14).

Los currículos analizados se centran en el desarrollo de competencias, pero la mayoría de tareas asignadas no involucra contextos reales y complejos (Loureiro, 2009: 101). En el discurso, los maestros dicen cumplir con lo establecido en el currículo, pero creen que son necesarias algunas adecuaciones para incluir conocimientos no explícitos en él. Además, aun cuando los maestros dicen cumplir con el currículo, se encontraron evidencias de que no siempre toman en cuenta metodologías y enfoques didácticos sugeridos en el mismo. (Loureiro, 2009, p. 103).

Otro aspecto del estudio se refiere al impacto que tienen sobre el trabajo docente las evaluaciones externas.

En unos países estudiados las evaluaciones nacionales proponen actividades significativas, con tareas de complejidad cognitiva que pocas veces se encuentra en las propuestas elaboradas por los maestros (Loureiro, 2009, p. 104). Esas evaluaciones en gran escala están más avanzadas y en línea con los desarrollos de la didáctica que las propuestas que se implementan en las aulas. Este hallazgo contradice la extendida idea de que las evaluaciones externas son memorísticas, están focalizadas en los resultados y desconocen los procesos de aprendizaje de los estudiantes, todo lo cual sólo sería contemplado en la evaluación que hacen los maestros en el aula. Si esto es verdad, las evaluaciones externas podrían apoyar la tarea diaria de los docentes, pero los efectos pueden ser ambivalentes: unos maestros pueden apropiarse las evaluaciones externas como herramientas, y aprovecharlas en función de su propia propuesta educativa, pero para otros pueden servir simplemente para “llenar el tiempo” y entrenar a los alumnos para responder pruebas estandarizadas. (Ravela, Picaroni y Loureiro, 2009, p. 11-12).

III. Estudios mexicanos

3.1 El trabajo de Vidales y Elizondo

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

La idea que tienen los maestros sobre la evaluación es fragmentada y dispersa, limitándose a reproducir la terminología de los cursos de actualización, por lo que las actividades de evaluación que llevan a cabo en el aula tienen cierto sustento teórico, pero no siempre bien entendido o utilizado en forma consciente. Por ello, para elaborar instrumentos o asignar calificaciones, 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. Algunas personas, sin embargo, piensan que no hay tal desconocimiento teórico, sino más bien resistencias debido al exceso de carga administrativa. (Vidales y Elizondo, 2005, p. 184)

En el discurso muchos maestros coinciden en que *se evalúa para mejorar la práctica docente y los desempeños de los alumnos en conocimientos, habilidades, actitudes y valores*. En la práctica prevalece la visión de que la evaluación es un evento temporal (bimestral, mensual o semanal) 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 y Elizondo, 2005, p. 185)

Según la opinión de supervisores y jefes de sector, un alto porcentaje de escuelas y maestros compran exámenes a editoriales y otras instancias. Para los exámenes bimestrales la cifra podría llegar hasta 80%. Sin embargo, una amplia mayoría de los directores (72%) sostiene que en sus escuelas los exámenes son hechos por los

maestros, y sólo 28% reconoce que los adquieren de instancias externas.

De los maestros que informan que elaboran sus propios exámenes, 66% dice hacerlo en forma individual y 34% en parejas o colegialmente. Sin embargo parece haber falta de capacitación en el manejo de los materiales curriculares en que se deberían basar las evaluaciones; pocos maestros los usan y muchos *no entienden sus propósitos o consideran que las formas de evaluación sugeridas son muy complicadas y laboriosas*; por ello los exámenes se suelen limitar a verificar cuántos datos aprendió el niño.

Según los maestros prevalece el uso de exámenes escritos y, al parecer, son los docentes de más antigüedad en el servicio los que menos utilizan esos recursos. Esto parecería indicio de que la falta de experiencia, o las deficiencias en la formación de las generaciones más recientes de maestros, podrían explicar que los maestros más jóvenes sean quienes más recurren a la adquisición de exámenes ya elaborados, dada su falta de preparación para hacerlos ellos mismos. (Vidales y Elizondo, 2005, p. 189-190)

En cuanto a la información derivada de las evaluaciones que se entrega a los padres, no hay informes amplios ni se da seguimiento a los resultados del proceso educativo; *sólo se entregan las boletas de calificación y las observaciones de apoyo a los hijos, de felicitación o de descalificación*. (Vidales y Elizondo, 2005, p. 194)

3.2 El estudio del Instituto Nacional para la Evaluación de la Educación

Entre 2009 y 2010 se diseñó y aplicó un cuestionario sobre sus propias prácticas de evaluación a 3534 profesores de 1193 escuelas de todo México, otro a 17908 alumnos y otro a los directores de los mismos planteles. La muestra de escuelas fue aleatoria y se calculó para que fuera representativa a nivel nacional.

Según el trabajo, la mayoría de docentes de primaria dicen utilizar la evaluación para mejorar el aprendizaje y la enseñanza, en forma congruente con un enfoque formativo. El 17% de los maestros encuestados afirmó que evalúa para saber cómo apoyar a los alumnos y 11% dijo hacerlo para planear y conducir sus clases. Un 43% dice emplear la evaluación sobre todo como medio para identificar problemas en el aprendizaje de los alumnos. (García Medina et al., 2011, p. 28)

La mayoría de los docentes (68%) aplican exámenes escritos bimestralmente (lo que coincide con el periodo establecido en la normativa para emitir calificaciones); 31% de los maestros lo hace mensual, quincenal o semanalmente y menos de 1% no los utiliza. Las preguntas orales son utilizadas frecuentemente por el 61% de los maestros. (García Medina et al., 2011, p. 71-72)

En español, la escritura de distintos textos y la lectura en voz alta son acciones de evaluación utilizadas con frecuencia por un mayor porcentaje de profesores: 67% dijo hacerlo en el bimestre. En cuanto a escritura, los maestros tienden a evaluar con mayor frecuencia la segmentación de las palabras y menos la claridad de la expresión de las ideas. Y en cuanto a lectura, el respeto de signos de puntuación y la dicción o pronunciación de las palabras son los aspectos que más profesores consideraron

importantes para valorar. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 44, 48-49).

En matemáticas sólo cuatro de cada 10 profesores utilizan problemas con mucha frecuencia para evaluar la asignatura. El 74% usa muy frecuentemente la actividad “resolver operaciones” para evaluar a sus alumnos. La actividad a la que recurren de manera más frecuente la mayoría de los docentes (82%) es resolver problemas en los que se tengan que hacer operaciones. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 49-50)

Por lo que se refiere a los niveles de demanda cognitiva que se manejan en las evaluaciones, algunos resultados del estudio del INEE van en el sentido de un uso considerable de actividades que implican niveles cognitivos altos. Los maestros mexicanos usarían con más frecuencia actividades de niveles altos e intermedios de complejidad, y con menor frecuencia las actividades de complejidad baja como memorización o razonamiento. Las primarias generales tuvieron porcentajes más bajos en las actividades de demanda cognitiva baja y mayores en casi todas las de nivel de complejidad alta. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 52 y 56).

Estos resultados no son consistentes con los que se refieren a la calificación de las evaluaciones, que apuntan en el sentido de que 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, Aguilera, Pérez y Muñoz, 2011, pp. 55 y 77).

Menos de 10% de los profesores utilizan ejemplos de malos trabajos para que los alumnos identifiquen lo que no deben hacer. La práctica más extendida (74%) es que los maestros indiquen a los alumnos las características requeridas para considerar de buena calidad un trabajo, pero esto no se acompaña por una práctica de modelamiento. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 72 y 75).

En cuanto a retroalimentación 70% de los profesores usan con frecuencia firmas o sellos, aunque no digan mucho al alumno. Una evaluación realmente formativa consiste en explicar cómo solucionar los errores cometidos, y algo más de 70% de los maestros dice utilizar esta estrategia de forma cotidiana. Aunque la cifra parece muy alta y su coincidencia con la anterior hace dudar de su validez, la opinión de los alumnos parece coincidir, pues cuando se les preguntó si sus maestros los apoyaban donde tenían más problemas para aprender, y si hablaban con ellos sobre lo que necesitarían hacer para mejorar su aprendizaje, en más del 70% de los grupos la mayoría de los estudiantes afirmó que sus profesores hacían de forma frecuente ambas actividades. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 76-78).

Lo anterior es poco congruente con otros resultados del estudio que indican que enseñar a los alumnos a autoevaluarse es actividad poco frecuente de la mayoría de los profesores: sólo 13% de los maestros suele pedir a sus alumnos que se autoevalúen y 46% nunca lo hace. Sólo 20% de los docentes incluye entre sus prácticas cotidianas dar guías o indicaciones para que los alumnos se autoevalúen o

coevalúen y casi 40% nunca lo hace. (García Medina, Aguilera, Pérez y Muñoz, 2011, pp. 78-79).

3.3 Diagnóstico de prácticas de evaluación de maestros de Nuevo León

El estudio se basó en una muestra representativa de primarias públicas, con 96 escuelas, 409 docentes (de 2°, 5° y 6° grado) y 3,358 alumnos de sexto. Se aplicó un cuestionario estructurado a los maestros y una entrevista semi-estructurada a uno por escuela. Se recogieron evidencias de evaluaciones de Matemáticas y Español. A los alumnos de sexto se aplicó además un cuestionario estructurado.

Casi todos los maestros dijeron haber hecho evaluación diagnóstica y siete de cada diez lo hizo en forma inmediata al inicio del ciclo escolar. Más de 90% dijo que para ese diagnóstico observaba trabajar a sus alumnos, aplicaba exámenes y revisaba antecedentes académicos; más de 80% dijo haber entrevistado a los alumnos o a sus padres; más de 70% dice que su planeación se basa en la información de esos ejercicios; y más de 50% refiere que entrega a cada alumno sus resultados haciendo recomendaciones.

Las prácticas más frecuentes, según los maestros, son dar retroalimentación individual a los alumnos (84%); involucrarlos en la autoevaluación de su avance (78%); y darles oportunidades para tomar decisiones y hacerse cargo de la forma en que aprenderán un tema (75%). Un número menor dice que suele fomentar que los alumnos discutan su avance en relación con objetivos centrales de aprendizaje (53%). Sólo 32% dice que limita el número de trabajos a los que asigna puntos para que los alumnos se concentren más en aprender que en la calificación.

En cuanto a los alumnos de menor rendimiento, más del 80% de los maestros dice que les indica las causas de sus errores al menos una vez por semana; 76% que les pide hacer actividades extra de refuerzo; 68% que les da materiales especiales para reforzar contenidos; 62% que lleva un seguimiento detallado de su avance; y 56% que dedica tiempo fuera de la clase para apoyar a esos alumnos.

Una proporción alta de maestros informa que realiza dos actividades diaria o casi diariamente para asignar calificaciones: revisar tareas y trabajos y observar el desempeño de los alumnos en el aula. Más del 80% dice hacer interrogatorios orales o pedir trabajos desde dos veces a la semana hasta una cada 15 días. Pero el instrumento más claramente asociado a la evaluación bimestral son las pruebas estructuradas que casi todos los maestros aplican: 82% de los alumnos dice haber presentado el examen bimestral; sólo 26% dice haber contestado exámenes *más o menos cada mes*; 21% cada semana y 11% *más o menos cada dos semanas*.

En la mayor parte de las escuelas los exámenes que se aplican son diseñados por los propios docentes, y los que dicen adquirirlos fuera mencionan las siguientes razones: 62% considera que son acordes a los programas de estudio, en tanto que 51% consideró que elaborar bien exámenes propios lleva mucho tiempo. Siete de cada diez maestros revisa y califica los exámenes que aplica; el resto favorece la autoevaluación

y la coevaluación.

Casi todos los maestros dicen usar todos los elementos mencionados para llegar a la calificación bimestral. El aspecto más importante según un mayor número de maestros es el desempeño en clase, seguido por los ejercicios en el cuaderno, el portafolio y los ejercicios de los libros de texto. Los resultados de exámenes, en cambio, ocupan el quinto lugar entre los puntos considerados para la calificación. Nueve de cada diez docentes dice informar a sus alumnos sobre los aprendizajes a evaluar en cada bimestre y los criterios de evaluación. Uno de cada tres dice dar a los padres un reporte escrito con información sobre el desempeño de sus hijos.

Según los alumnos, 45% de los maestros explican en detalle cómo van a calificar y 34% lo hacen en forma general. Dos de cada tres alumnos dicen que sus maestros brindan apoyo *bastantes veces o casi siempre* cuando tienen problemas para aprender, pero uno de cada tres señala que ese apoyo lo reciben *pocas veces o nunca*.

Las formas de apoyo más señaladas son: *trabaja con ellos a la hora del recreo o después de clases* (33%); *los pone a hacer trabajos extra o diferentes* (33%); y *los deja trabajando solos a la hora del recreo o después de clases* (26%). Poco más de una cuarta parte de los alumnos (28%) señala que *el maestro no hace nada en especial*. Según 57% de los alumnos, la actividad más frecuente de su maestro a fin de prepararlos para un examen es *indicarles temas a estudiar*; 38% dice que el maestro *indica páginas de los textos comprendidas en el examen*. Otras prácticas que implican más trabajo del profesor son menos usadas, según los alumnos: 42% darles un temario o cuestionario para el examen; 34% repasar con el grupo los temas que serán cubiertos; y 27% repasar las preguntas que vendrán.

La compra de exámenes bimestrales, según los docentes, ocurre sólo en una tercera parte de los casos y *la relación de las preguntas que vienen en el 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*; sólo 15% opina que los exámenes son en general ajenos a los temas vistos.

Un 78% de los alumnos considera que las preguntas *en su mayoría son fáciles, pero hay algunas difíciles* y 13% que todas las preguntas son fáciles. Los alumnos reportan como práctica frecuente que los exámenes son devueltos ya calificados (58%) y sólo 23% dice participar en la calificación. De las pruebas externas la gran mayoría de los alumnos dicen que son preparados para ellas por sus maestros: 15% con más de un mes de anticipación; 17% un mes antes; 31% dos semanas antes; y 31% una semana o menos antes. La preparación incluye *dedicar tiempo de algunas materias a repasar temas que vendrán en la prueba* (55%); *indicar temas a estudiar* (51%); y *resolver pruebas de años anteriores* (40%).

Casi todos los alumnos reportan la entrega de calificaciones bimestrales, variando la información adicional que dan los maestros. La forma más común de devolución es entregar calificaciones comentando razones (37%) y entregar calificaciones con los exámenes o trabajos (25%). Menos usual es dar calificación sin comentario (12%) y

con comentarios escritos (11%). En conjunto, seis de cada diez alumnos dicen recibir algún tipo de devolución además de su calificación.

En cuanto al contenido de la devolución que hacen los docentes, 50% de los alumnos informan que el maestro les dice en qué se deben fijar para ver si está bien o mal hecho su trabajo; 12% que les enseña otros exámenes o trabajos bien hechos para que comparen el suyo; 32% que platica sobre lo que necesitan hacer para mejorar su aprendizaje; 42% que da consejos sobre lo que deben hacer para mejorar; y 10% dice que el maestro no hace nada en especial.

IV. Conclusión

Stiggins y Conklin (1992, p. 4) afirman que los cimientos de las reformas educativas *serán débiles como arena si permanecemos mal informados sobre la naturaleza, el papel y la calidad de la evaluación en aula.*

Buenas investigaciones sobre las prácticas de evaluación serán el mejor punto de partida de esfuerzos para que los maestros desarrollen su competencia para hacer evaluaciones que apoyen el aprendizaje de sus alumnos. Esta revisión de literatura lleva a formular una hipótesis que podrá guiar trabajos posteriores.

La revisión de trabajos previos muestra que, por lo general, los maestros dicen estar de acuerdo con la evaluación formativa, pero varios elementos hacen temer que su práctica no sea congruente con sus concepciones y creencias:

- El 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 al parecer 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 incluso muy necesario recibir apoyo o formación para desarrollar mejor esta importantísima 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 y se confirmaría que cambiar las prácticas de evaluación supondrá esfuerzos prolongados y consistentes.

Referencias

Black, P. y William D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-75.

Briggs, D., Ruiz-Primo, A., Furtak, E., Shepard L. y Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.

Carter, K. (1984). Do teachers understand the principles for writing test? *Journal of Teacher Education*, 35(6), 57-60.

Dunn, K. y 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). Recuperado de <http://pareonline.net/getvn.asp?v=14&n=7>

Fleming, M. y 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* (pp. 29-38). San Francisco, CA: Jossey Bass.

García Medina, A., Aguilera, M. A., Pérez, M. G. y Muñoz, G. (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. y Nash, B. (2012). How many formative assessment angels can dance on the head of a meta-analytic pin. *Educational Measurement: Issues and Practice*, 31(4), 18-19.

Kingston, N. y Nash, B. (2011). Formative assessment: meta-analysis & call for research. *Educational Measurement: Issues & Practice*, 30(4), 28-37.

Loureiro, G. (2009). *Evaluación en el aula, currículo y evaluaciones externas*. Universidad Católica del Uruguay y GTEE-PREAL.

Martínez Rizo, F. (2012a). *La evaluación en el aula: promesas y desafíos de la evaluación formativa*. México: Universidad Autónoma de Aguascalientes.

Martínez Rizo, F. (2012b). La evaluación formativa del aprendizaje en el aula en la bibliografía en inglés y francés. Revisión de literatura. *Revista Mexicana de Investigación Educativa*, 17(54), 849-875.

Picaroni, B. (2009). *La evaluación en las aulas de primaria: usos formativos, calificaciones y comunicación con los padres*. Universidad Católica del Uruguay y Grupo de Trabajo sobre Estándares y Evaluación [GTEE-PREAL].

Ravela, P., Picaroni B. y Loureiro, G. (2009). *La evaluación de aprendizajes en las aulas de 6o. grado en América Latina*. Observatorio Regional de Políticas de Evaluación Educativa. Boletín No. 12, 1-12. GTEE-PREAL.

Reynolds, W. M. y Menard, K. A. (1980). *An investigation of teachers' test construction practices*. Ponencia presentada en la reunión anual del *National Council of Measurement in Education*. Boston, Massachusetts.

Ruiz, G. y Pérez, G. (en prensa). Creencias y prácticas de evaluación en aula de maestros de educación primaria de Nuevo León. En G. Ruiz (Ed.), *La evaluación en el aula: Diagnóstico* (Cap. 2). México: Universidad Autónoma de Aguascalientes.

Stiggins, R. J., Arter, J., Chappuis, J. y Chappuis, S. (2007). *Classroom assessment for student learning. Doing it right-using it well*. Upper Saddle River, NJ: Pearson.

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

Vidales, I. y Elizondo, M. D. (2005). *Prácticas de evaluación escolar en el nivel de educación primaria en el Estado de Nuevo León*. México: CAEIP/Santillana.