

EVIDENCIAS DE PRÁCTICAS DE EVALUACIÓN DE UN GRUPO DE PROFESORES DE PRIMARIAS DE NUEVO LEÓN

ADRIANA MERCADO SALAS / FELIPE MARTÍNEZ RIZO

Resumen:

Este artículo muestra los resultados del análisis de un conjunto de tareas asignadas y exámenes aplicados por una muestra no representativa de profesores de 2°, 5° y 6° grados de primarias del estado de Nuevo León. El trabajo partió de la revisión de investigaciones sobre las prácticas de evaluación de docentes a partir de la cual se elaboró un marco analítico con aspectos fundamentales de la evaluación formativa, como los referentes curriculares considerados, las instrucciones, la retroalimentación, los niveles de demanda cognitiva perseguidos y la forma de evaluación. El trabajo condujo a conclusiones tanto sustantivas como metodológicas. Se encontró que la práctica en el aula reflejada en las evidencias dista mucho de tener un enfoque formativo. Cabe señalar que la investigación se vio limitada al no tener información adicional sobre las tareas y exámenes recabados.

Abstract:

This article shows the results of analyzing a set of tasks assigned and examinations given by an unrepresentative sample of second-, fifth- and sixth-grade teachers in the state of Nuevo León. The project started with a review of research on teachers' evaluation practices, followed by the preparation of an analytical framework with fundamental aspects of formative evaluation, such as the curriculum referents considered, instructions, feedback, levels of cognitive demand, and the form of evaluation. The study led to substantive as well as methodological conclusions. The classroom practices reflected in the evidence were found to lack a formative focus. It should be pointed out that the research was limited by the absence of additional information about the compiled tasks and examinations.

Palabras clave: evaluación educativa, evaluación del aprendizaje, evaluación formativa, métodos de evaluación, educación básica, México.

Keywords: educational evaluation, evaluation of learning, formative evaluation, evaluation methods, elementary education, Mexico.

Adriana Mercado Salas es investigadora asociada y Felipe Martínez Rizo es investigador, ambos en el Departamento de Educación de la Universidad Autónoma de Aguascalientes. Av. Universidad 940, Ciudad Universitaria, 20131, Aguascalientes, Ags., México. CE: amercado@correo.uaa.mx / fmartin@correo.uaa.mx

Introducción

La importancia de la evaluación del aprendizaje de los alumnos ha llevado a desarrollar instrumentos que reflejen la situación en que se encuentra cada uno. Los más conocidos son las pruebas en gran escala –estandarizadas– cuyas consecuencias han sido, entre otras, el uso de los resultados para jerarquizar con fundamento poco sólido tanto a escuelas como a estudiantes y recompensar o castigar a los docentes. La conciencia de los límites de la información obtenida con las pruebas en gran escala para caracterizar las prácticas de los profesores ha hecho que se preste más atención a la evaluación realizada en aula por cada maestro, sea con exámenes parecidos a las pruebas de gran escala o mediante ejercicios, tareas, observaciones, trabajos grupales y otras actividades. Las evaluaciones a cargo del maestro son más frecuentes y tienen más impacto sobre el aprendizaje que las pruebas en gran escala, por lo que deberían recibir una atención al menos similar a la que se presta a estas últimas.

Este trabajo se deriva del estudio del diagnóstico de prácticas de evaluación de maestros que se hizo con una muestra representativa de primarias públicas del estado de Nuevo León (Ruiz Cuéllar, en prensa), en donde se aplicó un cuestionario estructurado a los docentes y se realizó una entrevista semiestructurada a un maestro por plantel; asimismo, se aplicó un cuestionario estructurado a los alumnos de sexto grado. Otro insumo obtenido en esta investigación fue un conjunto de fotografías de exámenes aplicados y tareas asignadas por una submuestra no representativa de maestros del estado de Nuevo León, sólo uno de cada tres profesores encuestados por escuela entregaba las evidencias fotográficas de manera libre.

Del análisis de las imágenes recabadas se esperaba captar aspectos de las prácticas de evaluación que ni los cuestionarios ni las entrevistas permitían detectar, y cuyos resultados mostramos en este texto.

Investigación sobre las prácticas de evaluación

Los estudios empíricos enfocados a explorar las prácticas de evaluación de los profesores son escasos, tal vez porque la complejidad de la docencia hace que en muchos casos no se incluyan entre los aspectos explorados. Esto llevó a realizar una revisión de literatura con resultados de trabajos empíricos sobre las prácticas de evaluación en aula de los maestros, los hallazgos fueron:

- 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 en México: Vidales y Elizondo (2005); García Medina *et al.* (2011); y Ruiz Cuéllar y Pérez Martínez (en prensa).

La obra *In Teachers' Hands*, de Stiggins y Conklin (1992), describe cómo evolucionaron los acercamientos metodológicos usados para analizar las prácticas de evaluación y los resultados obtenidos. El trabajo partió de una revisión de la literatura para dar paso a estudios empíricos que incluyeron una encuesta acompañada por un diario, observación participante, construcción de un marco analítico sobre el constructo, elaboración de perfiles detallados de prácticas de evaluación, análisis de evaluaciones aplicadas en el aula acompañadas de entrevistas con el objetivo de estudiar más a profundidad el nivel cognitivo de las habilidades fomentadas por los profesores y, finalmente, las prácticas de calificación de algunos maestros.

Por su parte, el proyecto latinoamericano fue realizado en ocho países e incluyó 80 escuelas y 160 maestros de sexto grado de primaria, a quienes se les aplicó un cuestionario y una entrevista. Además de lo anterior se tomaron registros fotográficos de propuestas de evaluación y se hizo un análisis de los currículos nacionales de los países participantes (Loureiro, 2009; Picaroni, 2009; Ravela, Picaroni y Loureiro 2009).

El trabajo de Vidales y Elizondo (2005) muestra los resultados de un proyecto elaborado en el estado de Nuevo León sobre prácticas de evaluación de docentes de primaria. En primer lugar se hizo un análisis documental seguido por entrevistas semiestructuradas con diferentes informantes (jefes de sector, directores, supervisores, etc.) y luego se trabajó con una muestra de docentes de primarias públicas.

Otro estudio mexicano fue el realizado entre 2009 y 2010 por el Instituto Nacional para la Evaluación de la Educación (García Medina *et al.*, 2011) en donde se diseñaron y aplicaron cuestionarios sobre prácticas de evaluación a una muestra nacional de profesores, sus alumnos y los directores de los mismos planteles.

El último estudio revisado, del que se deriva el presente artículo, es el diagnóstico de prácticas de evaluación de maestros realizado con una muestra representativa de primarias públicas del estado de Nuevo León (Ruiz Cuéllar, en prensa). Como mencionamos, se aplicó un cuestionario estructurado a maestros y otro a alumnos de sexto grado y se realizó una entrevista semiestructurada a un profesor por plantel.

La revisión de los 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. Por un lado, predomina una visión positiva sobre las competencias que se tienen para realizar evaluación formativa, pero la mayoría coincide en que es necesaria una capacitación o apoyo para desarrollar más esta competencia. Otro elemento que da cuenta de esta preocupación es que los maestros siguen dando gran importancia a la asignación de calificaciones a todos o casi todos los instrumentos (tareas y exámenes) de evaluación.

Otro hallazgo importante es que 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, lo 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.

Para contribuir a ampliar la información acerca de las prácticas de evaluación de los profesores, enseguida se presentan los resultados del análisis de un conjunto de evidencias (tareas y exámenes), con una explicación de la metodología utilizada y una síntesis de los hallazgos del análisis.

Metodología

El corpus de imágenes

Las imágenes recopiladas fueron mil 229, número que se redujo posteriormente a 850 imágenes analizables: 240 evidencias de español (inicialmente 317) y 213 de matemáticas (inicialmente 248). El 67% de las evidencias incluía una sola imagen; 24.1% tenía de 2 a 4 y 8.3% contaba con 5 o más. La depuración de las imágenes pasó por varias etapas, algunas se descartaron

porque no correspondían a tareas o exámenes o a las asignaturas de español o matemáticas, y otras por su baja calidad, demasiado borrosas para ser editadas. La última depuración se dio durante el análisis, al detectar evidencias que no contenían información para etiquetarlas como tareas o exámenes. Las imágenes analizadas fueron finalmente 850: 366 de matemáticas (213 evidencias de 74 escuelas) y 484 de español (240 evidencias de 73 escuelas).

El marco analítico

Para revisar las imágenes se preparó un marco de análisis que precisara las dimensiones de las prácticas evaluativas que idealmente se tratarían de detectar a partir tanto de la investigación sobre evaluación en aula en general y la formativa en particular como de la revisión de literatura adicional hecha para este proyecto y cuyos resultados se han reportado en el apartado anterior. Específicamente se consideraron el marco analítico desarrollado por Stiggins y Conklin (1992) y el utilizado en un estudio de validación de un instrumento complejo para medir las prácticas de evaluación en aula en el área de ciencias (Martínez *et al.*, 2012).

El primer marco comprende ocho dimensiones:

- 1) Propósitos de la evaluación
- 2) Métodos de evaluación: contenidos a evaluar y formas de hacerlo
- 3) Criterios para seleccionar el método de evaluación
- 4) Calidad de las evaluaciones
- 5) Retroalimentación
- 6) El maestro como evaluador
- 7) Percepción del estudiante acerca del maestro
- 8) Políticas de evaluación (Stiggins y Conklin, 1992:80).

El marco para validar el instrumento referido en el área de ciencias comprende nueve dimensiones que buscan captar el constructo *práctica de evaluación*:

- 1) Establecimiento de metas de aprendizaje claras
- 2) Frecuencia de las evaluaciones
- 3) Variedad de las evaluaciones
- 4) Alineación de las evaluaciones con las metas de aprendizaje
- 5) Complejidad cognitiva

- 6) Manejo de explicaciones y justificaciones científicas
- 7) Participación de los alumnos en actividades de auto-evaluación
- 8) Uso de la información para dar retroalimentación a los alumnos
- 9) Uso de la información para sustentar decisiones de instrucción (Martínez *et al.*, 2012:109).

A partir de lo anterior se elaboró una ficha para analizar las evidencias de evaluación que dieron los profesores de Nuevo León, con los siguientes aspectos:

- 1) Identificación de la evidencia:
 - a) Número de la escuela en que se produjo
 - b) Número de imágenes que corresponden a la evidencia
 - c) Grado escolar al que corresponde
 - d) Género del alumno que respondió el examen o hizo la tarea
 - e) Nivel de rendimiento del alumno, según su maestro
 - f) Medio en que se ubica la escuela (urbano o rural)
 - g) Asignatura: español o matemáticas
 - h) Tipo de evidencia: examen o tarea
 - i) *Origen de la evidencia: elaborada por el maestro, comprada...*
 - j) *Bloque del currículo al que corresponde*
- 2) Referentes curriculares
 - a) Presencia del referente curricular
 - b) *Referente enfatizado en español o matemáticas*
 - c) *Meta de aprendizaje del referente curricular*
 - d) Meta de aprendizaje prevaleciente en la evidencia (inferida)
 - e) Tipo de preguntas usadas (prevaleciente): de respuesta estructurada, de respuesta abierta corta, de respuesta extendida
 - f) Método de evaluación prevaleciente
 - g) *Coherencia entre meta del referente curricular y método prevaleciente*
 - h) Coherencia entre meta inferida y método prevaleciente
- 3) *Propósito de la evaluación*
 - a) *Propósito de uso de las evaluaciones de diagnóstico*
 - b) *Propósito de uso de evaluaciones bimestrales y tareas*
 - c) *Propósito de uso afectivo*

4) Instrucciones/Consigna

- a) Claridad de la instrucción
- b) Finalidad o propósitos
- c) Destinatarios
- d) Incertidumbre
- e) Restricciones
- f) Repertorio de metas
- g) Proceso de trabajo

5) Calificación

- a) Existencia de la calificación
- b) Claridad de los criterios de calificación
- c) *Ponderación de los criterios de calificación*

6) Retroalimentación

- a) Existencia de la retroalimentación
- b) Tipo de retroalimentación general
- c) *Calidad de la retroalimentación*
- d) Tipo de retroalimentación específica: valorativa, descriptiva u orientadora, devolutiva
- e) *Retroalimentación usada para la modificación de la práctica docente*

El análisis

Para cada aspecto se especificaron los valores que podrían estar presentes en las evidencias; la ficha incluyó códigos para facilitar la captura de los resultados.

Varias de las dimensiones anteriores se refieren a aspectos de las prácticas que no se pueden detectar mediante la sola revisión de las evidencias, y para hacer una mayor caracterización se necesitaría información adicional, como la relativa al propósito que perseguía el maestro con cierta evaluación o a forma de calificar. Por ello, para el análisis se elaboró una versión reducida de la ficha, con los aspectos que se podrían detectar con base en las evidencias. Los que se eliminaron fueron los que se destacan en cursivas en el listado anterior. De otras dimensiones, el análisis de las evidencias puede dar sólo información muy limitada, en particular todos los incisos del apartado 4 sobre instrucciones y consigna y los tres que no están en letra cursiva del numeral 6, sobre retroalimentación.

El trabajo fue hecho por cuatro analistas, previamente capacitados. Las evidencias a analizar se distribuyeron como se muestra en el cuadro 1.

CUADRO 1

Distribución de evidencias a calificar entre los cuatro analistas

Asignatura	Calificador	Evidencias	
Español	1	Exámenes	59
		Tareas	95
		Total	154
	2	Exámenes	58
		Tareas	95
		Total	153
	Evidencias analizadas por ambos calificadores		
Matemáticas	3	Exámenes	55
		Tareas	70
		Total	125
	4	Exámenes	53
		Tareas	70
		Total	123
	Evidencias analizadas por ambos calificadores		

Fuente: Elaboración propia a partir de la base de datos.

El análisis se llevó a cabo por dos parejas: una se ocupó de español y la otra de matemáticas. El conjunto de evidencias se distribuyó entre los integrantes de cada pareja, de tal manera que para español, 81 de las evidencias (de las que finalmente se pudieron correlacionar 67) fueron revisadas por dos personas; lo mismo ocurrió para matemáticas con 82 de las evidencias (de las que finalmente se recuperaron 72 para el análisis de consistencia). El resto de exámenes y tareas de ambas asignaturas únicamente fueron analizados por uno de los calificadores.

Lo anterior se hizo para verificar la consistencia de la calificación de cada analista según los diferentes aspectos considerados en el marco analítico. Para ello se calcularon los coeficientes de correlación entre calificadores.

El cuadro 2 muestra los resultados de un análisis elemental de consistencia entre calificadores. Las cifras resaltadas en negritas se refieren a los aspectos en los que hubo mayor consistencia, con cifras que rondan 0.7, en tanto que las resaltadas en cursivas son las más bajas, con cifras de menos de 0.1 a 0.3.

CUADRO 2

Consistencia de resultados entre calificadores

Dimensión	Aspectos	Total (n=139)	Español (n=67)	Matemáticas (n=72)
Referentes curriculares	Meta de aprendizaje	0.461	0.527	0.389
	Tipo de preguntas	0.547	0.740	0.345
	Método de evaluación	0.489	0.687	0.314
	Coherencia	<i>0.199</i>	<i>0.188</i>	<i>0.257</i>
Instrucción*	Claridad	0.658	0.682	0.635
	Finalidad	0.473	0.300	0.540
	Destinatarios	0.631	0.631	0.460
	Incertidumbre	0.658	0.677	0.613
	Restricciones	0.577	0.467	0.672
	Repertorio de metas	0.716	0.767	0.593
	Proceso de trabajo	<i>0.249</i>	<i>0.064</i>	<i>0.236</i>
Calificación	Existencia	0.224	0.464	0.305
Retroalimentación	Existencia	0.556	0.624	0.431
	Tipo general	0.595	0.619	0.568
	Valorativa	0.512	0.556	0.463
	Descriptiva	0.757	0.737	—
	Devolutiva	—	—	—
	Específica	0.665	0.644	0.695

* Para esta dimensión sólo se tomaron en cuenta las evidencias de tareas.

Fuente: Elaboración propia a partir de la base de datos.

En general, la consistencia entre los calificadores de las evidencias de español fue más alta que entre los de matemáticas. Lo anterior puede deberse a una menor claridad sobre los estándares usados en matemáticas, así como a que las evidencias en español, en general, muestran mayor información sobre la instrucción y la complejidad de las tareas.

Los aspectos que mostraron menor consistencia fueron los relacionados con la coherencia entre la meta y el método de evaluación inferido por los calificadores, el proceso de trabajo de las tareas y la existencia de la calificación, lo cual se debió, según los analistas a no tener más información acerca de las evidencias en estos aspectos.

Para rediseñar la ficha usada para el análisis de evidencias convendrá tomar como ejemplos las que ya se han utilizado para identificar palabras clave o elementos de los estándares de matemáticas que sean mejores indicadores de las dimensiones que se quiere estudiar.

Resultados

Los profesores que facilitaron las evidencias trabajaban mayoritariamente en escuelas urbanas (80.3%) y el resto en el medio rural. El 52.5% de las evidencias corresponde a español y 47.5% a matemáticas. Poco más de la mitad (51.2%) son tareas que los maestros asignan a los alumnos y 48.8% exámenes aplicados por los profesores. Del total, 173 corresponden a segundo grado, 113 a quinto y 167 a sexto de primaria. La proporción de trabajos de niños y niñas es cercana a la mitad, al igual que los alumnos que, a juicio de los profesores, tenían un nivel de rendimiento sobresaliente en el grupo.

La información obtenida con el análisis de las evidencias se muestra en tres apartados: sobre aspectos formativos de la práctica docente (el referente curricular, la instrucción, la calificación y la retroalimentación), sobre la autenticidad de las tareas, y sobre las metas de aprendizaje y los métodos de evaluación.

Aspectos formativos de la práctica docente

Este subapartado muestra los resultados del análisis de algunos aspectos del ejercicio docente relacionados con las prácticas formativas de evaluación. El análisis trató de recabar información de dicho proceso en las aulas, esto es, desde la elección de qué evaluar (referente curricular), cómo hacerlo (la instrucción dada y el tipo de preguntas usadas) y cómo comunicar los resultados a los alumnos (calificación y retroalimentación).

En el cuadro 3 se aprecia que ninguna evidencia de matemáticas tenía explícito el referente a evaluar. Algo similar se observó en español: muy pocas presentan el referente curricular, aunque un examen lo tenía, pero planteado confusamente. Algo mejor es la situación de las tareas: en 13 casos (11.4%) se encontró un referente, en 11 éste es confuso y en 2 explícito y claro. Lo anterior muestra que la mayoría de las evidencias (96.9%) no contienen los insumos para informar al alumno qué es lo que se pretende evaluar con las tareas y los exámenes aplicados y asignados en el salón de clase.

CUADRO 3

Caracterización del referente curricular de las evidencias

Evidencias	Caracterización del referente	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Explícito y claro	2	0.4	—	—	2	0.9
	Explícito y confuso	12	2.6	1	0.5	11	4.7
	No explícito	439	96.9	218	99.5	221	94.4
Español	Explícito y claro	2	0.8	—	—	2	1.8
	Explícito y confuso	12	5.0	1	0.8	11	9.6
	No explícito	226	94.2	125	99.2	101	88.6
Matemáticas	Explícito y claro	—	—	—	—	—	—
	Explícito y confuso	—	—	—	—	—	—
	No explícito	213	100	93	100	120	100

Fuente: Elaboración propia a partir de la base de datos.

En lo que se refiere a cómo evaluaron los profesores de la muestra, se encontró que hay pocas evidencias con instrucciones comprensibles y específicas (cuadro 4); con mayor frecuencia se encontró que son comprensibles no específicas, mientras que en una proporción considerable

éstas no existen o son incomprensibles. Esta situación es más marcada en las tareas, aunque no habría que perder de vista que el profesor puede dar instrucciones verbales o escribirlas en el pizarrón y por ello no aparecen en las evidencias. Por otro lado el maestro no necesita dar instrucciones para exámenes comprados pues éstos las incluyen.

CUADRO 4

Caracterización de las instrucciones que se dan a los alumnos

Evidencias	Caracterización de las instrucciones	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	No existen	156	35.1	23	10.5	133	58.8
	Incomprensibles	17	3.8	7	3.2	10	4.4
	Comprensibles sin especificar	238	53.5	163	74.4	75	33.2
	Comprensibles específicas	34	7.6	26	11.9	8	3.5
Español	No existen	88	36.7	13	10.3	75	65.8
	Incomprensibles	7	2.9	5	4.0	2	1.8
	Comprensibles sin especificar	142	59.2	106	84.1	36	31.6
	Comprensibles específicas	3	1.3	2	1.6	1	0.9
Matemáticas	No existen	68	33.2	10	10.8	58	51.8
	Incomprensibles	10	4.9	2	2.2	8	7.1
	Comprensibles sin especificar	96	46.8	57	61.3	39	34.8
	Comprensibles específicas	31	15.1	24	25.8	7	6.3

Fuente: Elaboración propia a partir de la base de datos.

Por otro lado, es evidente que la presencia de preguntas que requieran una respuesta extendida es muy reducida, en comparación con las que implican escoger entre opciones previamente estructuradas –mayoritarias en los exámenes– y las que suponen una respuesta corta, que prevalecen en las tareas (cuadro 5).

CUADRO 5
Caracterización de preguntas usadas en las evidencias

Evidencias	Caracterización de preguntas	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Respuesta estructurada	249	55.2	187	83.9	62	27.2
	Respuesta abierta corta	169	37.5	36	16.1	133	58.3
	Respuesta extendida	33	7.3	0	0	33	14.5
Español	Respuesta estructurada	144	60	115	91.3	29	25.4
	Respuesta abierta corta	64	26.7	11	8.7	53	46.5
	Respuesta extendida	32	13.3	0	0	32	28.1
Matemáticas	Respuesta estructurada	105	49.8	72	74.2	33	28.9
	Respuesta abierta corta	105	49.8	25	25.8	80	70.2
	Respuesta extendida	1	0.4	0	0	1	0.9

Fuente: Elaboración propia a partir de la base de datos.

Los cuadros 6, 7 y 8 precisan el tipo de respuesta que suponen las preguntas de cada uno de los tipos anteriores. Por lo que se refiere a las estructuradas, el cuadro 6 muestra que en los exámenes predominan abrumadoramente las de opción múltiple, mientras que en las tareas la mayoría requieren respuestas de completar, aunque también es importante la proporción de aquellas donde se deben relacionar columnas y de opción múltiple.

En cuanto a las preguntas que implican respuestas abiertas cortas, el cuadro 7 evidencia que en matemáticas casi todas admiten una sola respuesta correcta, tanto en exámenes como en tareas; mientras que en las tareas de español algo más de la mitad de las preguntas admiten más de una respuesta, e incluso en los exámenes esta situación está presente en más de una tercera parte de los casos.

En cuanto a preguntas que requieren una respuesta extendida, en el cuadro 8 puede apreciarse que prácticamente siempre están ausentes los criterios de evaluación y que, incluso, se encuentra algún caso en el que en realidad lo que se requiere es una respuesta corta.

CUADRO 6

Caracterización de las respuestas a preguntas de respuesta estructurada

Evidencias	Caracterización de respuestas	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Relacionar columnas	15	6.0	0	0	15	24.6
	Falso/verdadero	2	0.8	2	1.1	0	0
	Completar	47	19.0	19	10.2	28	45.9
	Opción múltiple	184	74.2	166	88.8	18	29.5
Español	Relacionar columnas	8	5.6	0	0	8	28.6
	Falso/verdadero	2	1.4	2	1.7	0	0
	Completar	22	15.4	11	9.6	11	39.3
	Opción múltiple	111	77.6	102	88.7	9	32.1
Matemáticas	Relacionar columnas	7	6.7	0	0	7	21.2
	Falso/verdadero	0	0	0	0	0	0
	Completar	25	23.8	8	11.1	17	51.5
	Opción múltiple	73	69.5	64	88.9	9	27.3

Fuente: Elaboración propia a partir de la base de datos.

CUADRO 7

Caracterización de las respuestas a preguntas de respuesta abierta corta

Evidencias	Caracterización del referente	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Implícita en pregunta	11	6.5	3	8.3	8	6.0
	Solo una respuesta	122	72.2	29	80.6	93	69.9
	Más de una respuesta	36	21.3	4	11.1	32	24.1
Español	Implícita en pregunta	9	14.1	2	18.2	7	13.2
	Solo una respuesta	21	32.8	5	45.5	16	30.2
	Más de una respuesta	34	53.1	4	36.4	30	56.6
Matemáticas	Implícita en pregunta	2	1.9	1	4.0	1	1.3
	Solo una respuesta	101	96.2	24	96	77	96.3
	Más de una respuesta	2	1.9	—	—	2	2.5

Fuente: Elaboración propia a partir de la base de datos.

CUADRO 8

Caracterización de las respuestas a preguntas de respuesta extendida

Evidencias	Caracterización del referente	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Con criterios de evaluación	—	—	—	—	—	—
	Sin criterios de evaluación	32	97	—	—	32	97
	Solo respuesta corta	1	3	—	—	1	3
Español	Con criterios de evaluación	—	—	—	—	—	—
	Sin criterios de evaluación	31	96.9	—	—	31	96.9
	Solo respuesta corta	1	3.1	—	—	1	3.1
Matemáticas	Con criterios de evaluación	—	—	—	—	—	—
	Sin criterios de evaluación	1	100	—	—	1	100
	Solo respuesta corta	—	—	—	—	—	—

Fuente: Elaboración propia a partir de la base de datos.

En cuanto a la comunicación de los resultados se encontró que una de las formas que típicamente se usa en las aulas es la calificación; sin embargo, el análisis de exámenes y tareas mostró que ésta no se encuentra en más de la mitad de los casos (55.4%), destacando el de las tareas de español (70%). De nuevo es posible que el profesor califique exámenes y/o tareas sin plasmar la calificación en las evidencias recopiladas, aunque parece menos probable que con las instrucciones. La misma observación se aplica a los criterios de calificación, ausentes en casi todos los casos (99.1%), lo que era esperable y muestra una limitación de la estrategia de obtención de información.

Con respecto a la retroalimentación que se ofrece a los alumnos, el cuadro 9 indica que no se encuentra en las evidencias revisadas en una proporción que va de 18 (en tareas de matemáticas) a 25% (en español) pero, nuevamente, es posible que los docentes retroalimenten sin que esto se plasme en evidencias textuales como las recogidas.

CUADRO 9

Caracterización de la retroalimentación que se ofrece en las evidencias

Evidencias	Caracterización de retroalimentación	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Existe	354	78.8	174	79.8	180	77.9
	No existe	95	21.2	44	20.2	51	22.1
Español	Existe	183	76.6	98	78.4	85	74.6
	No existe	56	23.4	27	21.6	29	25.4
Matemáticas	Existe	171	81.4	76	81.7	95	81.2
	No existe	39	18.6	17	18.3	22	18.8

Fuente: Elaboración propia a partir de la base de datos.

A pesar de que el porcentaje de evidencias sin retroalimentación es bajo (1 de cada 4), la que se encontró en su mayoría es de tipo valorativo en 98.6% (cuadro 10). Se observa una gran tendencia a señalar aciertos y errores,

marcar con algún sello, dar una calificación sin significado y apuntar en las tareas frases de aliento o desaliento a los alumnos (bien hecho, sigue así, esfuerzate más, etc.) (cuadro 11). Del total de las evidencias únicamente se encontraron 5 con retroalimentación descriptiva, en donde se explicaba al alumno el porqué de los aciertos o errores y el progreso que habían logrado.

CUADRO 10
Caracterización del tipo de retroalimentación

Evidencias	Tipo de retroalimentación	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Valorativa	345	98.6	169	99.4	176	97.8
	Descriptiva-orientadora	5	1.4	1	0.6	4	2.2
Español	Valorativa	176	97.2	95	99.0	81	95.3
	Descriptiva-orientadora	5	2.8	1	1.0	4	4.7
Matemáticas	Valorativa	169	100	74	100	95	100

Fuente: Elaboración propia a partir de la base de datos.

CUADRO 11
Caracterización del tipo de retroalimentación-valorativa

Evidencias	Retroalimentación valorativa	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Aciertos y errores	208	62.7	129	80.6	79	45.9
	Sellos-marcas-calif.	99	29.8	27	16.9	72	41.9
	Frases aliento-desaliento	25	7.5	4	2.5	21	12.2
Español	Aciertos y errores	84	51.5	60	69.8	24	31.2
	Sellos-marcas-calif.	71	43.6	26	30.2	45	58.4
	Frases aliento-desaliento	8	4.9	—	—	8	10.4
Matemáticas	Aciertos y errores	124	73.4	69	93.2	55	57.9
	Sellos-marcas-calif.	28	16.6	1	1.4	27	28.4
	Frases aliento-desaliento	17	10.1	4	5.4	13	13.7

Fuente: Elaboración propia a partir de la base de datos.

Autenticidad de las tareas y demanda cognitiva de las evaluaciones

El análisis de las evidencias disponibles comprendió un elemento más que amplía el anterior, aunque únicamente en relación con las tareas, tanto de español como de matemáticas. Si bien éstas se prestan, más que los exámenes, para pedir la realización de actividades de mayor complejidad, que supongan más tiempo y comprendan varios aspectos, esto no parece ser frecuente de acuerdo con el análisis, según se podrá ver más adelante. El currículo de la primaria mexicana considera este tipo de tareas complejas con el concepto de proyectos en el campo formativo de lengua y comunicación. Un concepto que apunta en la misma dirección es el de tareas auténticas.

Una tarea auténtica se asemeja a un problema real, en contraposición con una artificial, que sólo puede darse en un contexto escolar. De acuerdo con la conceptualización de Wiggins, los rasgos que distinguen las tareas auténticas incluyen un propósito claro, se ubican en un contexto real, con destinatarios más allá del maestro, presentan elementos de incertidumbre, tienen restricciones y más de una solución aceptable, implican un proceso de trabajo largo (con momentos de retroalimentación a productos parciales), y para su solución es necesario poner en juego habilidades de diversa complejidad, ya que incluyen un repertorio variado de metas de aprendizaje, también de diferente nivel (Wiggins, 1998:23-24).

El cuadro 12 presenta los resultados del análisis relativo a la presencia de rasgos de autenticidad en las tareas de español y matemáticas revisadas.

Salta a la vista que las tareas analizadas tienen muy pocos rasgos de autenticidad. En la mayoría de los casos el propósito no es explícito, y en casi todos los restantes simplemente se menciona, solamente se describe en 4 casos, 2 en cada asignatura analizada. El contexto casi nunca es real: es escolar una gran proporción en matemáticas, mientras en español lo es en dos terceras partes de los casos y semi-escolar en la restante. Las tareas presentan poca incertidumbre y, en mayor medida, únicamente admiten una solución. Coincidiendo con lo ya observado anteriormente, implican sobre todo mecanización en el caso de matemáticas, y mecanización o conocimiento en español. Metas de nivel alto son escasas en español y prácticamente inexistentes en matemáticas.

CUADRO 12

Caracterización de la autenticidad de las tareas (n = 226)

Aspecto	Caracterización	Total	%	Español	%	Matemát.	%
Finalidad o propósito	No existe	150	66.4	84	74.3	66	58.4
	Se menciona	72	31.9	27	23.9	45	39.8
	Se describe	4	1.8	2	1.8	2	1.8
Contexto	Escolar	174	76.7	73	64.6	101	88.6
	Semi-escolar	51	22.5	38	33.6	13	11.4
	Real	2	0.9	2	1.8	—	—
Incertidumbre	Incluye datos necesarios	149	65.9	54	47.8	95	84.1
	No incluye todos los datos	47	20.8	33	29.2	14	12.4
	Enfrenta al alumno	30	13.3	26	23	4	3.5
Restricciones	Una solución	167	73.9	72	63.7	95	84.1
	Alternativas de solución	59	26.1	41	36.3	18	15.9
Repertorio de metas de aprendizaje	Mecanización	120	52.9	43	38.1	77	67.5
	Conocimiento-Razonamiento	92	40.5	56	49.6	36	31.6
	Desempeño-Producción	15	6.6	14	12.4	1	0.9

Fuente: Elaboración propia a partir de la base de datos.

A partir de los cuadros anteriores es posible inferir, con fundamento razonable, que la mayoría de las preguntas de las evidencias revisadas, especialmente las de los exámenes, difícilmente podrán referirse a metas de aprendizaje que impliquen una demanda cognitiva alta, ya que se prestan más para evaluar habilidades de baja complejidad, como memorización y mecanización.

También parece razonable la presencia un poco menos marcada de preguntas de respuesta estructurada en las tareas para que los alumnos hagan en casa o en el aula misma, en particular aquellas de opción múltiple, falso/verdadero, relacionar columnas y completar; especialmente si se considera la alta dificultad de diseñar tareas que incluyan preguntas cerradas de ese tipo.

El cuadro 13 aporta información que complementa lo anterior (tomando en cuenta tanto tareas como exámenes), al presentar la cantidad de evidencias cuya realización supone un proceso relativamente breve y, presumiblemente, simple, puesto que deben hacerse en una sola ocasión con un tiempo limitado. La presencia de evidencias que supongan un proceso más prolongado, con diversos momentos, es completamente marginal, incluso sin el requerimiento adicional de entregas parciales. En forma esperable, la situación es un poco menos marcada en español y en las tareas.

CUADRO 13

Caracterización del proceso de trabajo implicado

Evidencias	Caracterización del proceso	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Una ocasión con tiempo limitado	383	97	218	99.1	165	94.3
	Diferentes momentos sin entregas parciales	12	3.0	2	0.9	10	5.7
Español	Una ocasión con tiempo limitado	231	96.3	126	100	105	92.1
	Diferentes momentos sin entregas parciales	9	3.8	0	0	9	7.9
Matemáticas	Una ocasión con tiempo limitado	152	98.1	92	97.9	60	98.4
	Diferentes momentos sin entregas parciales	3	1.9	2	2.1	1	1.6

Fuente: Elaboración propia a partir de la base de datos.

Llegamos así al punto más importante del análisis: la complejidad de la meta de aprendizaje a la que se refieren las evidencias, en términos de la demanda cognitiva que implica responder las preguntas correspondientes. Debe reiterarse que se trata de la complejidad de la meta de aprendizaje *inferida* a partir de la revisión de las evidencias, y no de la que pudo tener en mente el docente al diseñar o seleccionar la actividad de evaluación de que se trate. El cuadro 14 presenta las conclusiones del análisis de las evidencias revisadas en este sentido.

En una gran parte las evidencias tienen un nivel de complejidad bajo: las tareas y los exámenes a los alumnos demandan el ejercicio de actividades simples o repetitivas que, en el esquema utilizado, corresponden a las categorías “mecanización” y “conocimiento”. Lo anterior se presenta tanto en español como en matemáticas. Sólo se pide a los alumnos acciones como nombrar, enlistar, elegir y seleccionar. En matemáticas la proporción de evidencias que plantean únicamente actividades de mecanización ronda los dos tercios, tanto en exámenes como en tareas; la parte restante consiste en actividades algo más complejas, de conocimiento. El caso de las tareas de español es el único en que cerca de 50% de las evidencias piden el nivel de conocimiento, donde se espera que los alumnos describan o expliquen algún tema.

CUADRO 14

Caracterización de la complejidad cognitiva de la meta de aprendizaje inferida

Evidencias	Caracterización de la meta	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Mecanización	261	57.6	143	64.1	118	51.3
	Conocimiento	166	36.6	75	33.6	91	39.6
	Razonamiento	18	4.0	5	2.2	13	5.7
	Habilidad de desempeño	2	0.4	—	—	2	0.9
	Habilidad p/crear productos	6	1.3	—	—	6	2.6
Español	Mecanización	123	51.3	84	66.7	39	34.2
	Conocimiento	93	38.8	38	30.2	55	48.2
	Razonamiento	17	7.1	4	3.2	13	11.4
	Habilidad de desempeño	1	0.4	—	—	1	0.9
	Habilidad p/crear productos	6	2.5	—	—	6	5.3
Matemáticas	Mecanización	138	64.8	59	60.8	79	68.1
	Conocimiento	73	34.3	37	38.1	36	31
	Razonamiento	1	0.5	1	1.0	—	—
	Habilidad de desempeño	1	0.5	—	—	1	0.9
	Habilidad p/crear productos	—	—	—	—	—	—

Fuente: Elaboración propia a partir de la base de datos.

De las 453 evidencias solamente se hallaron 18 (17 de español y 1 de matemáticas) donde los niveles de complejidad demandados llegaron a ser de razonamiento; 2 tareas más, una de cada asignatura, requerían habilidades de desempeño y 6 tareas de español, habilidades para crear productos. En síntesis, sólo 4% de las tareas y exámenes revisados piden a los alumnos realizar actividades en donde lleven a cabo análisis, síntesis, deducción o justificación.

CUADRO 15

Caracterización de la complejidad cognitiva de la meta de aprendizaje por grados

Grado	Evidencias	Caracterización de la meta	Total		Exámenes		Tareas	
			Número	%	Número	%	Número	%
2°	Total	Mecanización	132	76.3	67	76.1	65	76.5
		Conocimiento	40	23.1	21	23.9	19	22.4
		Razonamiento	1	0.6	—	—	1	1.2
	Español	Mecanización	60	65.9	37	75.5	23	54.8
		Conocimiento	30	33.0	12	24.5	18	42.9
		Razonamiento	1	1.1	—	—	1	2.4
	Matemáticas	Mecanización	72	87.8	30	76.9	42	97.7
		Conocimiento	10	12.2	9	23.1	1	2.3
	5°	Total	Mecanización	48	42.5	21	44.7	27
Conocimiento			57	50.4	24	51.1	33	50.1
Razonamiento			5	4.4	2	4.3	3	4.5
Habilidad p/crear productos			3	2.7	—	—	3	4.5
Español		Mecanización	18	31.0	11	39.3	7	23.3
		Conocimiento	32	55.2	15	53.6	17	56.7
		Razonamiento	5	8.6	2	7.1	3	10.0
		Habilidad p/crear productos	3	5.2	—	—	3	10.0
Matemáticas		Mecanización	30	54.5	10	52.6	20	55.6
		Conocimiento	25	45.5	9	47.4	16	44.4

(CONTINÚA)

Grado	Evidencias	Caracterización de la meta	Total		Exámenes		Tareas	
			Número	%	Número	%	Número	%
6°	Total	Mecanización	81	48.5	55	62.5	26	32.9
		Conocimiento	69	41.3	30	34.1	39	49.4
		Razonamiento	12	7.2	3	3.4	9	11.4
		Habilidad de desempeño	2	1.2	—	—	2	2.5
		Habilidad p/crear productos	3	1.8	—	—	3	3.8
	Español	Mecanización	45	49.5	36	73.5	9	21.4
		Conocimiento	31	34.1	11	22.4	20	47.6
		Razonamiento	11	12.1	2	4.1	9	21.4
		Habilidad de desempeño	1	1.1	—	—	1	2.4
		Habilidad p/crear productos	3	3.3	—	—	3	7.1
	Matemáticas	Mecanización	36	47.4	19	48.7	17	45.9
		Conocimiento	38	50.0	19	48.7	19	51.4
		Razonamiento	1	1.3	1	2.6	—	—
		Habilidad de desempeño	1	1.3	—	—	1	2.7

Fuente: Elaboración propia a partir de la base de datos

Por lo que se refiere a exámenes, en el cuadro 15 se puede apreciar que, para el segundo grado, todas las preguntas para ambas asignaturas sólo implicaban mecanización o conocimiento. En quinto, la proporción siguió siendo de 100% en matemáticas y en español bajó a 93%; mientras en sexto fue de 97 y 96%, respectivamente; es decir, en los tres grados prácticamente la totalidad de los exámenes analizados correspondían a estos niveles. En cuanto a las tareas, en matemáticas la proporción de preguntas de baja demanda cognitiva fue de 100% tanto en segundo como en quinto grados y de 97% para sexto. Una situación mejor aparece en el caso de las tareas de español, campo formativo en el que en segundo grado las preguntas de baja demanda cognitiva representaron 97%, proporción que bajó a 80% en quinto y a 69% en sexto.

Si bien mecanización y conocimiento son actividades necesarias como base para el desarrollo de otras más complejas –y es normal que su presencia sea mayoritaria en los primeros grados del trayecto escolar– la permanencia

casi exclusiva de preguntas que no pasan de esos niveles en los últimos grados de la primaria, en especial en matemáticas, parece preocupante.

Metas de aprendizaje y métodos de evaluación.

Congruencia entre enseñanza y evaluación

Tras todo lo anterior es posible un último análisis de las evidencias disponibles, que se refiere directamente a la hipótesis a la que se llegó al final del diagnóstico de las prácticas de evaluación de maestros de primaria de Nuevo León. En tal sentido, parecería haber congruencia entre las prácticas de evaluación y las de enseñanza de los docentes, pero en tanto unas y otras se refieren a propósitos de aprendizaje de bajo nivel de complejidad o de demanda cognitiva.

De tal modo, se volvieron a caracterizar todas las evidencias, siguiendo ahora la tipología de métodos de evaluación de Stiggins *et al.* (2007:100), con categorías ligeramente diferentes a las utilizadas antes sobre el tipo de preguntas y respuestas que se mostraron en los cuadros 6 a 8. Allí se pudo ver que en poco más de la mitad de los casos se usan métodos de selección de respuestas (respuesta estructurada). La nueva caracterización confirmó que 8 de cada 10 exámenes utilizan esta modalidad, mientras el resto sólo permite al alumno dar una respuesta corta única. También se confirma que esta última es más usada en las tareas, en comparación con los exámenes (cuadro 16).

Finalmente se realizó un último análisis a partir de un complemento del marco utilizado, y que se puede ver en el anexo (cuadro A.2), basado en Stiggins *et al.* (2007); se trata de un cuadro de dos dimensiones: cada columna se refiere a uno de cuatro métodos de evaluación (preguntas de respuesta estructurada, preguntas de respuesta abierta, evaluaciones de desempeño o ejecución y preguntas formuladas oralmente, en una interacción del maestro con los alumnos), y cada fila a uno de cinco tipos de metas de aprendizaje (mecanización, conocimiento, razonamiento, habilidad de desempeño y habilidad de crear productos). En cada casilla del cuadro se señala qué tan adecuado es, en principio, el método de la columna con la meta del renglón correspondiente. Se precisa “en principio” porque teóricamente es posible pensar en buenas preguntas de cualquier tipo para evaluar metas de aprendizaje de cualquier nivel, pero en la práctica unos métodos son sin duda más apropiados para metas de cierto nivel. Los resultados de este último análisis se muestran en el cuadro 17.

CUADRO 16

Caracterización del método de evaluación prevaleciente

Evidencias	Caracterización del referente	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Totales	Selección de respuesta	236	52.4	181	81.5	55	24.1
	Respuesta abierta corta	201	44.7	41	18.5	160	70.2
	Evaluación de desempeño	13	2.9	—	—	13	5.7
	Comunicación personal	—	—	—	—	—	—
Español	Selección de respuesta	132	55.2	109	87.2	23	20.2
	Respuesta abierta corta	94	39.3	16	12.8	78	68.4
	Evaluación de desempeño	13	5.4	—	—	13	11.4
	Comunicación personal	—	—	—	—	—	—
Matemáticas	Selección de respuesta	104	49.3	72	74.2	32	28.1
	Respuesta abierta corta	107	50.7	25	25.8	82	71.9
	Evaluación de desempeño	—	—	—	—	—	—

Fuente: Elaboración propia a partir de la base de datos.

En los exámenes el método de evaluación usado en la mayoría de las evidencias es una opción muy buena en relación con el nivel cognitivo de la meta de aprendizaje: en 87.2% de los exámenes de español y en 75.3% de los de matemáticas se da esa combinación. En español, el resto se divide por igual en opciones buenas y aceptables, y en matemáticas el segundo lugar lo ocupan las aceptables. En ninguna de las asignaturas hay casos de malas opciones, pero dado que se trata de metas de aprendizaje inferidas esto parece lógico. También puede explicarse porque en los hechos hay congruencia entre enseñanza y evaluación, con énfasis en metas de aprendizaje de bajo nivel cognitivo, según la hipótesis apuntada. En las tareas la situación es algo diferente, pero nuevamente en español las opciones muy buena y buena suman 82.5% y en matemáticas esas dos opciones acumulan 52.7% de los casos, con 46.5% más de opciones aceptables.

CUADRO 17
*Caracterización de la coherencia entre meta
 de aprendizaje y método de evaluación*

Evidencias	Caracterización de la coherencia	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
Total	Muy buena opción	267	59.3	182	82.0	85	37.3
	Buena opción	80	17.8	11	5.0	69	30.3
	Opción aceptable	96	21.3	29	13.1	67	29.4
	Mala opción	7	1.6	—	—	7	3.1
Español	Muy buena opción	162	67.8	109	87.2	53	46.5
	Buena opción	49	20.5	8	6.4	41	36
	Opción aceptable	22	9.2	8	6.4	14	12.3
	Mala opción	6	2.5	0	0	6	5.3
Matemáticas	Muy buena opción	105	49.8	73	75.3	32	28.1
	Buena opción	31	14.7	3	3.1	28	24.6
	Opción aceptable	74	35.1	21	21.6	53	46.5
	Mala opción	1	0.5	—	—	1	0.9

Fuente: Elaboración propia a partir de la base de datos.

En síntesis, la congruencia entre métodos de evaluación y tipo de metas de aprendizaje en cuanto a nivel de demanda cognitiva es alta. Los elementos anteriores del análisis que se ha presentado permiten afirmar que la congruencia se da porque tanto las metas de aprendizaje como los métodos de evaluación se refieren a actividades que suponen bajos niveles de demanda cognitiva.

Conclusiones

En lo sustantivo, con las limitaciones a las que se hará referencia, podemos afirmar que las prácticas de evaluación de los profesores estudiados distan mucho de tener un componente formativo importante. Las opiniones más

optimistas que los docentes mismos expresan en cuestionarios y entrevistas no deben entenderse, sin embargo, como intentos conscientes por dar información falsa, sino como un reflejo de comprensiones superficiales del sentido real de la evaluación formativa.

El hallazgo más importante al que llegamos al explorar las prácticas en una forma que ofrece una visión algo más amplia que los cuestionarios y las entrevistas aplicados a los maestros, encuentra en particular que los profesores estudiados utilizan en forma masiva exámenes integrados básicamente por preguntas de respuestas estructuradas, que implican operaciones de baja demanda cognitiva, y que asignan tareas que en general, aunque en un grado algo menor al de los exámenes, implican también solamente memorización o repeticiones mecánicas. Asimismo, encontramos un grado de congruencia bastante alto entre los métodos de evaluación utilizados y las metas de aprendizaje inferidas, que en ambos casos corresponden a niveles de demanda cognitiva bajos. Otros hallazgos particulares, que resultan lógicos a partir de lo anterior, son que:

- Las consignas se reducen por lo general a indicaciones elementales sobre la forma de responder, pero carecen por completo, en casi todos los casos, de rasgos de autenticidad.
- La calificación de exámenes y tareas privilegia el conteo de aciertos y errores, sin referencia a estándares de desempeño precisos.
- La retroalimentación se suele reducir a informar acerca del resultado de la calificación, en la forma de una nota a la que se llegó mediante un proceso que la hace carecer de significado pedagógico, que frecuentemente va acompañada de elogios o amonestaciones, así como de exhortaciones a mejorar sin orientación sobre cómo conseguirlo.

La afirmación de que estos hallazgos particulares resultan lógicos se fundamenta señalando que cuando la meta de aprendizaje consiste en que el alumno memorice una serie de datos o realice mecánicamente una serie de operaciones matemáticas, entonces es razonable que la consigna sea muy sencilla, sin rasgos de autenticidad, que la calificación se limite a contar aciertos y errores, y que la retroalimentación se reduzca a informar al estudiante de lo anterior, diciéndole que es necesario que avance más cuando el número de aciertos esté lejos del total posible.

Por otra parte, es obvio que el proyecto no permitió analizar la forma en que los docentes estudiados llevan a cabo actividades de evaluación cotidianamente, observando el desempeño de sus alumnos, interrogándolos o poniéndolos a hacer ejercicios cortos, en forma individual o grupal, ni tampoco sobre la manera de dar retroalimentación verbal o de promover la autoevaluación de cada alumno sobre su propio avance, o la co-evaluación entre unos y otros. El acercamiento utilizado no fue suficiente para ello, lo que nos lleva al segundo tipo de conclusiones.

Las limitaciones a las que alude el párrafo anterior se refieren a que la recolección de evidencias consistió únicamente, como se ha explicado ya, en pedir a maestros de Nuevo León que permitieran fotografiar algunos ejemplos de exámenes aplicados por ellos y de tareas asignadas a sus alumnos; sin embargo, no se les interrogó sobre la manera en que los elaboraron o seleccionaron, ni acerca de la consigna que presentaron a los alumnos o de la forma de calificar o retroalimentar a los estudiantes. Esta fue, obviamente, una deficiencia que deberá corregirse en trabajos posteriores, acompañando la recolección de evidencias de evaluación con instrumentos apropiados para explorar las dimensiones de las prácticas que no se pueden detectar simplemente con el análisis de los documentos, sino que requieren información adicional que es necesario solicitar expresamente al maestro.

Anexos

CUADRO A1

Metas de aprendizaje

Tipo de meta	Explicación	Estándares de contenido/ palabras clave
Mecanización	Se deben aprender algunos conocimientos /hechos/ conceptos momentánea y memorísticamente En matemáticas elaboración de ejercicios repetitivos, sumas, restas, multiplicación, etc.	Nombrar, listar, recordar, memorizar, elegir, seleccionar

(CONTINÚA)

Tipo de meta	Explicación	Estándares de contenido/ palabras clave
Conocimiento	<p>Se deben aprender algunos conocimientos/hechos/conceptos por completo; algunos son recuperados utilizando materiales de referencia.</p> <p>En matemáticas se requiere recordar y llevar acabo fórmulas como simple ejercicio</p>	<p>Explicar, comprender, describir, identificar, informar, definir, etiquetar, asociar, reconocer</p>
Razonamiento	<p>Habilidades del pensamiento- utilizando el conocimiento para resolver problemas, tomar decisiones, planear, etc.</p>	<p>Analizar: componentes, partes, secuencia lógica, pasos, la idea principal, detalles de apoyo, determinar, seccionar, examinar, ordenar.</p> <p>Comparar/contrastar: discriminar entre iguales y diferentes, distinguir entre las similitudes y diferencias, yuxtaponer (asociar o juntar).</p> <p>Sintetizar: combinar, formular, organizar, adaptar, modificar.</p> <p>Clasificar: categorizar, ordenar, agrupar, dar ejemplos</p> <p>Inferir/deducir: interpretar, identificar implicaciones, predecir conclusiones, crear hipótesis, generalizar.</p> <p>Evaluar: justificar, fundamentar una opinión, pensar críticamente, valorar, criticar, debatir, defender, juzgar, probar</p>
Habilidades de desempeño	<p>Demostración observable, en donde lo que importa es lo que se hace; se utiliza el conocimiento y el razonamiento para realizar una tarea</p>	<p>Observar, enfocar la atención, escuchar, realizar, hacer, preguntar, dirigir, trabajar, leer, hablar, ensamblar, operar, utilizar, demostrar, medir, investigar, guiar, simular, recoger, dramatizar, explorar</p>
Habilidad para crear productos	<p>Las características del producto final es lo importante; se utiliza el conocimiento, el razonamiento y las habilidades de desempeño para producir el producto final</p>	<p>Diseñar, crear, desarrollar, producir, escribir, representar, mostrar, crear un modelo, construir</p>

CUADRO A2

Meta a ser evaluada	Método de evaluación			
	<i>Selección de respuesta</i>	<i>Respuesta abierta</i>	<i>Evaluación de desempeño</i>	<i>Interacción verbal</i>
Mecanización	(++) Opción muy buena si sólo se pretende corroborar memorización y mecanización	(+-) Opción aceptable cuando se trata de un grupo de estudiantes, pero tiende a consumir tiempo	(-) Mala opción: consume demasiado tiempo para evaluar todos los temas previstos	(+-) Opción aceptable con grupos de estudiantes, pero es una opción que tiende a consumir tiempo
Conocimiento	(++) Muy buena opción para evaluar el dominio de metas de nivel conocimiento	(+) Opción buena si se trata de verificar la comprensión de las relaciones entre los elementos del conocimiento	(-) Mala opción: consume demasiado tiempo para cubrir todo	(+-) Opción aceptable evaluando las respuestas, e infiriendo el dominio, pero es una opción que tiende a consumir tiempo
Razonamiento	(+-) Es una opción aceptable únicamente en la evaluación de la comprensión de algunos patrones de razonamiento	(++) Muy buena opción: la descripción escrita de soluciones a problemas complejos puede dar información para evaluar diversos aspectos de razonamiento	(++) Muy buena opción, pues permite observar a los alumnos resolver algunos problemas e inferir su nivel de razonamiento	(++) Muy buena opción: permite pedir al estudiante que "piense en voz alta" o que responda a preguntas de seguimiento para sondear su razonamiento
Habilidades de desempeño o ejecución	(-) Mala opción. Permite evaluar el dominio de los prerrequisitos del conocimiento necesarios para un desempeño competente, pero no se puede basar sólo en ellos para evaluar la habilidad misma		(+) Buena opción. Se pueden observar y evaluar las habilidades a medida que se realizan	(++) Muy buena opción cuando la habilidad es la comunicación oral; si se trata de otras habilidades no es buena
Habilidad de crear productos	(-) Mala opción. Permite evaluar conocimientos que son prerrequisitos de la habilidad de crear productos de calidad, pero no se puede utilizar para la evaluación de la calidad de los productos mismos	(++) Muy buena opción cuando el producto es escrito. No es buena cuando el producto no es escrito	(++) Muy buena opción. Puede evaluar los atributos del producto mismo	(-) Mala opción

Referencias

- García Medina, Adán Moisés; Aguilera García, María Antonieta; Pérez Martínez, María Guadalupe y Muñoz Abundez, Gustavo (2011). *Evaluación de los aprendizajes en el aula. Opiniones y prácticas de docentes de primaria en México*, Ciudad de México: Instituto Nacional para la Evaluación de la Educación.
- Loureiro Denis, Graciela (2009). *Evaluación en el aula, currículo y evaluaciones externas*, s.l.: Instituto de Evaluación Educativa/UCU/GTEE-PREAL.
- Martínez Rizo, Felipe (2012). *La evaluación en el aula: promesas y desafíos de la evaluación formativa*, Aguascalientes: Universidad Autónoma de Aguascalientes.
- Martínez, José Felipe; Borko, Hilda; Stecher, Brian; Luskin, Rebecca y Kloser, Matt (2012). "Measuring classroom assessment practice using instructional artifacts: A validation study of the QAS Notebook", *Educational Assessment*, vol. 17, núms. 2-3, pp. 107-131.
- Picaroni, Beatriz (2009). *La evaluación en las aulas de primaria: usos formativos, calificaciones y comunicación con los padres*, s.l.: Instituto de Evaluación Educativa/UCU/GTEE-PREAL.
- Ravela, Pedro; Picaroni, Beatriz y Loureiro, Graciela (2009). "La evaluación de aprendizajes en las aulas de 6° grado en América Latina", *Boletín del Observatorio Regional de Políticas de Evaluación Educativa* (GTEE-PREAL), núm. 12 (julio), pp. 1-12.
- Ruiz Cuéllar, Guadalupe (ed.) (en prensa). *La evaluación en el aula: diagnóstico*, Aguascalientes: Universidad Autónoma de Aguascalientes.
- Ruiz, Cuéllar Guadalupe y Pérez Martínez, Guadalupe (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*, Aguascalientes: Universidad Autónoma de Aguascalientes.
- Stiggins, Richard J. y Nancy F. Conklin (1992). *In teacher's hands: Investigating the practices of classroom assessment*, Albany: State University of New York Press.
- Stiggins, R.; Arter, J.; Chappuis, J. y Chappuis, S. (2007). *Classroom assessment for student learning. Doing It right-using It well*, Portland: ETS.
- Vidales D., Ismael y Elizondo G., María D. (2005). *Prácticas de evaluación escolar en el nivel de educación primaria en el estado de Nuevo León*, Ciudad de México: CAEIP/Santillana.
- Wiggins, Grant (1998). "Ensuring authentic performance", en *Educative assessment: Designing assessments to inform and improve student performance*, San Francisco: Jossey-Bass, pp. 21-42.

Artículo recibido: 12 de junio de 2013

Dictaminado: 16 de octubre de 2013

Segunda versión: 2 de diciembre de 2013

Aceptado: 2 de diciembre de 2013