

## METAS DE APRENDIZAJE Y MÉTODOS DE EVALUACIÓN EN LAS AULAS DE PRIMARIAS DE NUEVO LEÓN

ADRIANA MERCADO SALAS / FELIPE MARTÍNEZ RIZO  
Universidad Autónoma de Aguascalientes

**RESUMEN:** El estudio sobre las prácticas de evaluación que los profesores llevan a cabo en aula ha sido en los últimos años un esfuerzo por conocer, comprender y mejorar la calidad educativa que se lleva a cabo en las escuelas. En este sentido, esta investigación trató de dar cuenta sobre uno de los muchos aspectos a considerar en tan complejo fenómeno educativo: las tareas asignadas y los exámenes aplicados por una muestra de profesores de primaria del estado de Nuevo León. El enfoque particular del estudio fue conocer sobre aspectos que la RIEB ha considerado como primordiales en la práctica docente, esto es, la evaluación formativa en el aula. El trabajo sintetiza los resultados del análisis hecho a un conjunto de evidencias de tareas y exámenes de las asignaturas de español y matemáticas partiendo de la elaboración de un marco analítico que

consideraba lo siguiente: los referentes curriculares evaluados en las evidencias, las instrucciones, la retroalimentación y en especial los niveles de demanda cognitiva perseguidos en estos artefactos (metas de aprendizaje) y sus métodos de evaluación. Si bien el trabajo nos llevó a conclusiones tanto sustantivas como metodológicas una de las más importantes fue fundamentar y corroborar la hipótesis derivada del diagnóstico elaborado por Ruiz Cuéllar y Pérez (en prensa) en donde concluían que tanto las prácticas de evaluación y las de enseñanza de los docentes tenían una gran congruencia, pero ambas en bajos niveles de demanda cognitiva.

**PALABRAS CLAVE:** Evaluación Educativa, Evaluación del aprendizaje, Evaluación formativa, Métodos de evaluación, Educación básica.

### Introducción

La evaluación del aprendizaje de los alumnos se realiza de dos formas: en gran escala, con pruebas estandarizadas, o en aula, por cada maestro en donde algunos de los instrumentos utilizados son exámenes parecidos a las pruebas de gran escala, además de los anteriores se evalúa 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. Con esta idea se han llevado a cabo una serie de investigaciones en la línea de la evaluación Formativa en aula

en la Universidad Autónoma de Aguascalientes, en dónde en un primer momento comenzó con la revisión de una extensa bibliografía y la elaboración de un estado del conocimiento (Martínez Rizo, 2012). El objetivo central fue hacer un diagnóstico de las prácticas de evaluación de los maestros de primaria de Nuevo León, con base en la información obtenida mediante cuestionarios que se aplicaron a maestros y alumnos, y en entrevistas hechas a los docentes (Ruiz Cuéllar y Pérez Martínez, en prensa). Esta fase incluyó la recolección de evidencias fotográficas de exámenes y tareas hechas por los alumnos, de cuyo análisis se esperaba captar aspectos de las prácticas de evaluación que cuestionarios y entrevistas no permiten detectar. En seguida se presentan los resultados del análisis de esas evidencias, con una explicación de la metodología utilizada y una síntesis de los hallazgos del análisis.

## Metodología

### El corpus de imágenes

Las imágenes recopiladas fueron 1229, número que se redujo luego a 850 imágenes analizables, 366 de matemáticas (217 evidencias de 74 escuelas) y 484 de español (240 evidencias de 73 escuelas). La depuración de las imágenes pasó por varias etapas en las que se descartaron algunas, por no corresponder a tareas o exámenes o a las asignaturas de español o matemáticas o por ser de baja calidad, demasiado borrosas para ser editadas.

### El marco analítico

Para el análisis de las imágenes se preparó un marco analítico que precisara las dimensiones de las prácticas de evaluación que idealmente se trataría de detectar, a partir de la literatura sobre evaluación en aula en general, y en particular sobre evaluación formativa (Martínez Rizo, 2012), así como de la revisión de literatura adicional hecha para el mismo. Se consideró en particular el marco analítico desarrollado por Stiggins y colaboradores (1992), así como el marco utilizado en un estudio de validación de un instrumento complejo para medir las prácticas de evaluación en aula en el área de ciencias (*QAS Notebook*, Martínez et al., 2012)

La versión de la ficha de análisis comprendió los siguientes aspectos:

- A. Identificación de la evidencia: número de la escuela en la que se produjo, número de imágenes que correspondían a la evidencia, grado escolar, género del alumno

que respondió la evidencia, nivel de rendimiento del alumno, según su maestro, medio, asignatura, tipo de evidencia.

- B. Referentes curriculares: presencia, meta de aprendizaje prevaleciente (inferida), preguntas usadas (prevaleciente), método de evaluación prevaleciente, coherencia entre meta inferida y método prevaleciente.
- C. Instrucciones/Consigna: claridad, finalidad, destinatarios, incertidumbre, restricciones, repertorio de metas, proceso de trabajo.
- D. Calificación: existencia, claridad de los criterios de calificación.
- E. Retroalimentación: existencia, retroalimentación general, retroalimentación específica.

### El análisis

Para cada aspecto se especificaron valores que podrían estar presentes además de sufrir cambios durante el análisis. Varias de las dimensiones anteriores y otras detectadas se refieren a aspectos de las prácticas que no se pueden detectar mediante la sola revisión de las evidencias, sino que para caracterizarlos bien se necesitaría información adicional, como la relativa al propósito que perseguía el maestro con cierta evaluación, o a la forma de calificar.

El análisis se llevó a cabo por dos parejas: una pareja llevo a cabo el análisis de evidencias de español, mientras a la otra correspondió el análisis de las evidencias de matemáticas, de tal manera que 163 evidencias fueron revisadas por dos personas (81 de español y 82 de matemáticas), el resto de las evidencias de ambas asignaturas solo fueron analizadas por uno de los calificadores. Lo anterior se hizo para poder verificar la consistencia de la calificación de cada pareja según los diferentes aspectos. Para ello se calcularon los coeficientes de correlación entre calificadores mediante un análisis elemental de consistencia entre calificadores.

Los aspectos en los que hubo mayor consistencia entre calificadores, con cifras que rondan el 0.7 fueron los relacionados con algunos rasgos de autenticidad (instrucción/consigna). En tanto que los aspectos con consistencia más baja rondan cifras de menos de 0.1 a 0.3, entre estos aspectos están el proceso de trabajo de las tareas y la existencia de la calificación, lo cual se debió, según los calificadores a no tener más información acerca de las evidencias en estos aspectos.

En general la consistencia entre los calificadores de las evidencias de español fue más alta que entre los calificadores de evidencias de matemáticas.

## Resultados

Ninguna evidencia de matemáticas tenía explícito el referente a evaluar. Algo similar se observó en español: muy pocas presentaron el referente curricular (14), 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 este es confuso y en dos explícito y claro.

Hay pocas evidencias con instrucciones comprensibles y específicas (7.6%). La situación más frecuente fue la de instrucciones comprensibles no específicas (53.5%). En el 39 % de las evidencias las instrucciones no existen o son incomprensibles (39%), situación más usual en las tareas.

En cuanto a calificación, en más de la mitad de los casos no se encuentra (55.4%), destacando el caso de las tareas de español, en que la cifra llega al 70%. Por otro lado los criterios de calificación estuvieron ausentes en casi todos los casos (99.1%).

Por lo que se refiere a la retroalimentación que se ofrece a los alumnos, no se encuentra en las evidencias revisadas en una proporción que va del 18 por ciento (en tareas de matemáticas) al 25 por ciento (en español). Lo anterior no parece ser preocupante hasta el momento en que el análisis fue especificando la naturaleza de la retroalimentación en donde se encontró que mayoritariamente (98.6%) son de tipo valorativas, esto es, contabilizar aciertos y errores o poner sellos o frases de aliento o desaliento. Sólo cinco evidencias muestran una retroalimentación descriptiva en donde se explicaba al alumno el porqué de los aciertos o errores y el progreso que habían logrado. No habría que perder de vista que el profesor puede dar instrucciones, calificaciones y retroalimentación de forma verbal o escrita en el pizarrón, y por ello éstas no aparecen en las evidencias.

Es evidente que la presencia de preguntas que requieran una respuesta extendida es muy reducida (7.3%), en comparación con las que implican escoger entre opciones

estructuradas (55.2%), que son mayoritarias en los exámenes, y las que suponen respuesta corta, que prevalecen en tareas.

A partir de lo encontrado anteriormente se puede inferir que la mayoría de las preguntas de las evidencias revisadas, especialmente las de los exámenes, difícilmente podían referirse a metas de aprendizaje que implicarán una demanda cognitiva alta, ya que se prestan más para evaluar habilidades de baja complejidad, como memorización y mecanización.

Información adicional que complementa lo anterior es la cantidad de evidencias cuya realización supone un proceso relativamente breve (97%) y, presumiblemente simple, puesto que deben hacerse en una sola ocasión con un tiempo limitado.

Llegamos así al punto más importante del análisis: el que se refiere a la complejidad de la meta de aprendizaje a la que se refieren las evidencias. Debe reiterarse que se trata de la complejidad de la meta de aprendizaje *inferida* a partir de la revisión de las mismas, y no de la que pudo tener en mente el docente al diseñar o seleccionar la actividad de evaluación de que se trate.

En una gran mayoría, las evidencias tienen un nivel de complejidad bajo (94%): las tareas asignadas y los exámenes aplicados demandan el ejercicio de actividades simples o repetitivas, que en el esquema utilizado corresponden a las categorías “mecanización” y “conocimiento”.

De las 457 evidencias sólo se hallaron 18 (17 de español, una de matemáticas) en los que los niveles de complejidad demandados llegaron a ser de razonamiento; dos tareas más, una de español y otra de matemáticas, requerían habilidades de desempeño y seis tareas de español habilidades para crear productos. En síntesis, sólo cuatro por ciento de las tareas y exámenes revisados piden a los alumnos realizar actividades en donde lleven a cabo análisis, síntesis, deducción o justificación.

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.

El análisis de las evidencias disponibles comprendió un aspecto más, que amplía el anterior, en relación únicamente con las tareas, tanto de español como de matemáticas. Las tareas se prestan más que los exámenes para que se pida la realización de actividades de mayor complejidad, que supongan más tiempo y comprendan varios aspectos. El currículo de la primaria mexicana contempla 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 a una tarea artificial, que sólo puede darse en un contexto escolar. De acuerdo con la conceptualización de Wiggins (1998), los rasgos que distinguen las tareas auténticas incluyen que tienen un propósito claro, se ubican en un contexto real, con destinatarios más allá del maestro, presentan elementos de incertidumbre y tienen restricciones, pero tienen más de una solución aceptable, y para su solución es necesario poner en juego habilidades de diversa complejidad.

Es notorio que las tareas analizadas tienen pocos rasgos de autenticidad. En la mayoría de los casos el propósito no es explícito (66.4%), y en casi todos los restantes simplemente se menciona (31.9%) y sólo se describe en cuatro casos, dos en español y dos en matemáticas. El contexto casi nunca es real: es masivamente escolar en matemáticas (88.6%), y en español lo es también en dos terceras partes de los casos, y semi-escolar en la tercera parte restante. Las tareas presentan poca incertidumbre, y mayoritariamente sólo admiten una solución (alrededor del 70%).

### **Congruencia entre enseñanza y evaluación**

Por último fue posible un último análisis, partiendo de la hipótesis propuesta al final del diagnóstico de las prácticas de evaluación de maestros de primaria de Nuevo León (Ruiz y Pérez, en prensa) en el sentido de que parecería haber congruencia entre las prácticas de evaluación y las de enseñanza de los docentes, pero ambas se refieren bajos niveles de demanda cognitiva.

Se volvieron a caracterizar todas las evidencias, siguiendo la tipología de métodos de evaluación de Stiggins et al. (2007: 100). La nueva caracterización confirmó lo encontrado respecto a los tipos de respuesta en las evidencias en donde ocho de cada 10 exámenes usan métodos de selección de respuesta (estructurada), mientras el resto sólo permiten al alumno dar una respuesta corta única. También se confirma que el método de respuesta corta es más usado en las tareas que los profesores asignan, en comparación con los exámenes.

A partir de esta nueva caracterización y de un complemento del marco analítico utilizado basado en Stiggins se pudo identificar el nivel de coherencia que existe entre el referente inferido y el método usado para evaluarlo. En principio, cada método propuesto por Stiggins corresponde de mejor manera a cierto tipo de meta de evaluación. 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 la Tabla 1.

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 opciones aceptables. En ningún caso existen coherencias negativas, pero dado que se trata de metas de aprendizaje inferidas esto parece lógico. También puede explicarse porque en los hechos haya 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.

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

Hay conclusiones sustantivas, sobre las prácticas de evaluación mismas, y otras metodológicas, sobre el acercamiento al estudio de un fenómeno tan complejo.

En lo sustantivo podemos afirmar que las prácticas de evaluación de los profesores de Nuevo León distan de tener un componente formativo importante. El hallazgo más importante al que llegamos al explorar las prácticas de esta forma es que en particular los maestros utilizan en forma masiva exámenes integrados básicamente por preguntas de respuesta estructurada, que implican operaciones de baja demanda cognitiva. Además, encontramos una gran congruencia entre los métodos de evaluación utilizados y las metas de aprendizaje perseguidas (inferidas), que en ambos casos corresponden a niveles de demanda cognitiva bajos.

Otros hallazgos importantes 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; y, la retroalimentación se suele reducir a informar sobre el resultado de la calificación, en la forma de una nota así como de exhortaciones a mejorar sin orientación. La afirmación de que estos hallazgos particulares resultan lógicos se fundamenta en la idea de que cuando la meta de aprendizaje consiste en que el alumno memorice, o bien 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.

Por otra parte, es obvio que el proyecto no permitió estudiar la forma en que los docentes llevan a cabo actividades de evaluación cotidianamente, el acercamiento utilizado no fue suficiente para ello, lo que nos lleva al segundo tipo de conclusiones. Las limitaciones se refieren al hecho de que la recolección de evidencias consistió únicamente en pedir a los maestros permitieran fotografiar algunos ejemplos de exámenes aplicados

por ellos y de tareas asignadas a sus alumnos, pero no se les interrogó sobre aspectos como: cómo seleccionaron dichos exámenes y tareas, la consigna que presentaron, la forma de calificar o la de retroalimentar. Esta fue, obviamente, una deficiencia que deberá corregirse en trabajos posteriores recabando información adicional a las evidencias y que es necesario solicitar expresamente a los profesores.

## Tablas y figuras

**Tabla 1. Coherencia entre meta de aprendizaje y método de evaluación**

Evidencias	Coherencia	Total		Exámenes		Tareas	
		Número	%	Número	%	Número	%
<b>Total</b>	Muy buena opción	267	59.3	182	82	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	0	0	7	3.1

Fuente: Elaboración propia a partir de la base de datos

## Referencias

Martínez, José Felipe et al. (2012). Measuring Classroom Assessment Practice Using Instructional Artifacts: A Validation Study of the QAS Notebook. *Educational Assessment*, Vol. 17, N° 2-3, pp. 107-131.

Martínez Rizo, Felipe (2012). *La evaluación en el aula: Promesas y desafíos de la evaluación formativa*. Aguascalientes: Universidad Autónoma de Aguascalientes.

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

Stiggins, Richard J. y Nancy F. Conklin, (1992). *In teacher's hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.

Wiggins, Grant (1998). Ensuring Authentic Performance. En *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass. Cap. 2, pp. 21-42.