



## DISEÑO Y VALIDACIÓN DE UN PROTOCOLO DE OBSERVACIÓN PARA EVALUAR LAS ACTIVIDADES DE ENSEÑANZA EN MATEMÁTICAS (POAEM)

**Lesly Yahaira Rodríguez Martínez**  
Centro de Investigación y Docencia Económicas

**María Guadalupe Pérez Martínez**  
CONACYT - Universidad Autónoma de Aguascalientes

---

**Área temática:** Evaluación Educativa.

**Línea temática:** Diseño y validación de instrumentos.

**Tipo de ponencia:** Reporte final de investigación.

---

### ***Resumen:***

El propósito de este estudio fue diseñar un Protocolo de Observación de las Actividades de Enseñanza en Matemáticas (POAEM), que permitiera recabar información válida y confiable sobre las actividades de enseñanza en quinto grado de primaria. Los participantes fueron 20 docentes de escuelas públicas, a quienes se videograbó en dos ocasiones. Se recabaron 40 videograbaciones, seccionadas en 222 tareas matemáticas, las cuales fueron la unidad de análisis de las actividades de enseñanza.

Los resultados mostraron que la estructura dimensional del POAEM se agrupa en un factor y no en los cuatro factores deducidos a partir de la revisión de literatura. En el estudio de generalizabilidad se encontró que las dimensiones del POAEM en conjunto explican cerca del 80% de la varianza.

Se espera que el proceso seguido para desarrollar este estudio sea un referente a nivel nacional para futuras investigaciones que tengan como propósito diseñar instrumentos para estudiar las prácticas docentes, pues éstas son fundamentales para la mejora de la calidad de la educación en cualquier sistema educativo.

***Palabras claves:*** Actividades de enseñanza, Educación primaria, Protocolo de observación, Estudio metodológico, POAEM.

## Introducción

En México, las últimas reformas educativas han enfatizado la importancia de favorecer un modelo de enseñanza encaminado a promover un aprendizaje que priorice la participación y reflexión continua de los estudiantes a través del diálogo, la colaboración y la construcción de conocimientos (SEP, 2011; SEP, 2017). Esto involucra una forma de enseñanza cuyos principios didácticos se fundamenten en lo que el conocimiento científico ha encontrado que puede favorecer un aprendizaje profundo, es decir, a través de una enseñanza que retome las características diferenciadas del alumnado; promueva el descubrimiento del propio conocimiento y favorezca un aprendizaje significativo (McTighe y Wiggins, 2012; Aubuson *et al.*, 2014).

El modelo de enseñanza actual sugiere el desarrollo de actividades de enseñanza que impliquen desafíos intelectuales para los estudiantes, así como actividades que resulten relevantes de acuerdo con sus contextos. Por ejemplo, actividades con propósitos definidos, que establezcan diferentes niveles de exigencia cognitiva, promuevan el uso de diferentes herramientas de apoyo para el aprendizaje, impulsen la construcción de nuevos conocimientos, y permitan la implementación de diversas estrategias de solución, entre otros (SEP, 2011; SEP, 2017).

Autores como Newmann, Marks y Gamoran (1996), Newmann, López y Bryk (1998) y Newmann, Bryk y Nagaoka, (2001) sugieren considerar tres criterios para valorar la calidad de la actividad de enseñanza a partir del trabajo intelectual que se promueve con los estudiantes: a) *investigación disciplinada*, que se refiere a la puesta en práctica de conocimientos previos y procedimentales para desarrollar nuevos conocimientos; b) *construcción del conocimiento*, que implica producir el conocimiento y lograr un significado más que una memorización; y c) *valor más allá del contexto escolar*, que tiene que ver con el significado de los aprendizajes para los estudiantes en contextos diferentes al escolar.

Los estudios realizados por Newmann y colaboradores, han sido un parteaguas para investigaciones que toman las actividades de enseñanza como un potencializador de oportunidades de aprendizaje. Estudios al respecto han adoptado diferentes enfoques de investigación, por ejemplo, la intervención (Avery y Freeman, 2002), el diseño de nuevos modelos de enseñanza (McTighe, Seif y Wiggins, 2004; Wiggins y McTighe, 2008), y estudios metodológicos (Boston y Wolf, 2006; Aubusson, Burke, Shuck, Kearney y Frischknecht, 2014).

En México el acercamiento a las prácticas de enseñanza en matemáticas se ha realizado en su mayoría a través de la caracterización de las prácticas de enseñanza, las creencias y concepciones de los docentes, el análisis del dominio que tienen los docentes de los contenidos, y el aprovechamiento de los estudiantes. Por lo que resulta necesario impulsar una línea de trabajo centrada en el desarrollo de instrumentos que permitan obtener información de mayor calidad a través de estudios metodológicos que generen diferentes evidencias de validez y confiabilidad (Ávila *et al.*, 2013; Martínez y Chávez, 2016).

Considerando lo anterior, el objetivo de este estudio fue diseñar un Protocolo de Observación para Evaluar las Actividades de Enseñanza en quinto grado de primaria en la asignatura de Matemáticas (POAEM) y

obtener evidencias de validez y confiabilidad de la información recabada con el instrumento diseñado. Es un estudio metodológico cuyo aporte es el desarrollo de un instrumento para estudiar un fenómeno educativo, así como sustentar juicios de valor que se establezcan a partir de la realidad observada con base en criterios e indicadores específicos. Las preguntas de investigación se centraron en aspectos propios de la información recabada a partir del protocolo de observación diseñado:

1. ¿Hasta qué punto la información recabada con el protocolo de observación parece ser una medida válida?
2. ¿Cuál es la confiabilidad de las medidas de observación de las actividades de enseñanza a partir del protocolo de observación diseñado?
3. ¿Cuál es la estructura factorial del protocolo de observación para evaluar las actividades de enseñanza en la asignatura de matemáticas?
4. ¿Hay evidencias de que las medidas de observación de las actividades de enseñanza en matemáticas en quinto grado de primaria son consistentes entre los observadores?
5. ¿A qué se debe la variabilidad de los datos obtenidos con el protocolo de observación diseñado? Por ejemplo, a los jueces, ocasiones de observación, docentes, contexto escolar, entre otras facetas de error.

Los alcances de este estudio no permiten reportar información sobre el tipo de actividades de enseñanza que implementan los docentes en la asignatura de matemáticas; sin embargo, se espera que su diseño y validación sea un insumo para estudios futuros interesados en diferentes aspectos de las actividades de enseñanza en matemáticas.

## Desarrollo

A través de la enseñanza se promueven los aprendizajes, esto hace que sea un indicador de las oportunidades de aprendizaje que los docentes ofrecen a los estudiantes durante las sesiones de clase, ya que las acciones implementadas por los docentes en las aulas influyen en la construcción de significados de los estudiantes (Boston y Wolf, 2006).

Llinares (2008), Verschaffel, Greer y De Corte (2010), Boston y Wolf, (2006) y Aubusson, Burke, Shuck y Kearney (2014) sugieren favorecer acciones a través de actividades de enseñanza que promuevan el desarrollo de habilidades del pensamiento y aplicar procesos de aprendizaje con flexibilidad y creatividad. Esto con el objetivo de establecer condiciones propicias para que los estudiantes tengan oportunidades para comprometerse con el razonamiento, logren sus propios significados y desarrollen una comprensión profunda.

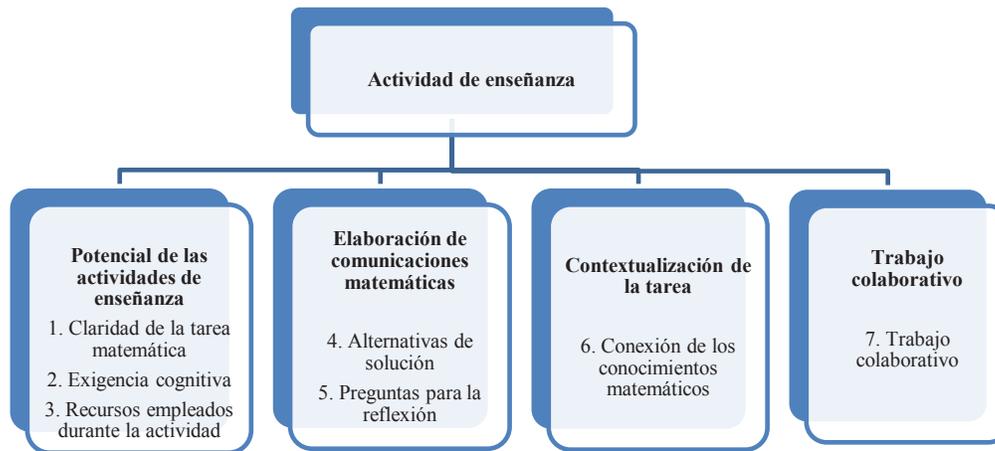
En la literatura se han identificado aspectos que coinciden con las condiciones mencionadas. Por ejemplo, el National Council of Teachers of Mathematics (NCTM) señala que “el aprendizaje de las matemáticas depende fundamentalmente de lo acontecido en el salón de clases” (2015, p. 8); esto es, del tipo de trabajo intelectual propuesto por los docentes a través de las actividades de enseñanza y de las oportunidades que éstas representan para la construcción de nuevos conocimientos, para estimular el razonamiento e involucrar a los estudiantes de manera significativa con las matemáticas (Picaroni y Loureiro, 2010).

Algunas características que se deben considerar para el desarrollo de las actividades de enseñanza en matemáticas incluyen: a) metas o propósitos de aprendizaje que guíen dichas actividades; b) diferentes niveles de exigencia cognitiva; c) construcción de nuevos conocimientos; d) diversidad de estrategias de solución; e) preguntas que fortalezcan el aprendizaje; f) cercanía de las actividades con el contexto de los estudiantes; g) utilizar representaciones matemáticas; y f) promover el trabajo colaborativo (Llinares, 2008; NCTM, 2015, y Swan, 2015).

Modelos de enseñanza como la Pedagogía Auténtica, la Enseñanza para la Comprensión y el Modelo de Entendimiento por Diseño, sugieren características clave como: a) comunicar claramente los propósitos de aprendizaje; b) ofrecer oportunidades para enfrentar diferentes niveles de exigencia cognitiva; c) promover la construcción de conocimientos a partir de los contenidos o conocimientos previos; d) promover alternativas de solución; e) contextualizar las actividades de enseñanza para que los estudiantes puedan darles sentido a partir de sus propias experiencias; f) promover una investigación disciplinada; g) favorecer el trabajo colaborativo, así como la comunicación de resultados, y otorgar tiempos para una autoevaluación por parte de los estudiantes (Avery y Freeman, 2002).

Tomando como referencia las características identificadas durante la revisión de literatura, el modelo teórico para el diseño del POAEM se basa en siete dimensiones para valorar el constructo “actividades de enseñanza”. La selección de las dimensiones se realizó con base en las coincidencias en los criterios propuestos en diferentes modelos de enseñanza para las matemáticas y en general. La siguiente figura resume las dimensiones del modelo teórico del instrumento diseñado.

Figura 1: Modelo teórico para evaluar las actividades de enseñanza en matemáticas



Fuente: Diseño propio a partir de la revisión de literatura.

Para evaluar las siete dimensiones del POAEM se desarrollaron rúbricas con cuatro niveles de desempeño en las que se incluyen criterios e indicadores específicos:

- **Claridad de la tarea matemática:** identifica la manera en que el docente comunica las tareas matemáticas a través de los productos o aprendizajes esperados, el lenguaje matemático, y estrategias para favorecer la comprensión de los estudiantes sobre el desarrollo de la tarea propuesta.
- **Exigencia cognitiva:** incluye indicadores para identificar tareas promotoras de un aprendizaje superficial (conocimientos aislados o de repetición sin sentido para los estudiantes), así como aquellas que favorecen un aprendizaje profundo (tareas que requieren analizar, explicar, justificar procesos, etc.).
- **Recursos empleados:** se centra en el uso que el docente y estudiantes hacen de artefactos y materiales disponibles para convertirlos en recursos que apoyen el aprendizaje de las matemáticas. Los indicadores tienen que ver con la manera en que el recurso se relaciona con el contenido trabajado.
- **Alternativas de solución:** incluye indicadores para identificar las oportunidades que el docente ofrece a los estudiantes para emplear diferentes estrategias o procedimientos de solución, así como para identificar las pistas o sugerencias que pueden sesgar las respuestas de los estudiantes.
- **Preguntas para la reflexión:** se focaliza en tres indicadores: a) el tipo de preguntas que requieren que los estudiantes describan información; b) preguntas que ayudan a los estudiantes

a hacer evidente su pensamiento matemático; y c) preguntas que requieren que los estudiantes justifiquen y demuestren el trabajo realizado.

- **Conexión de los conocimientos matemáticos:** basado en la clasificación de contextos de tareas matemáticas propuesta por Picaroni y Loureiro (2010), por lo que se busca identificar experiencias que los estudiantes puedan reconocer por su situación o contexto, la conexión de las situaciones problemáticas con aspectos del mundo que nos rodea; y la relación de ideas o conceptos matemáticos con otros temas o disciplinas.
- **Trabajo colaborativo:** incluye indicadores relacionados con el papel del docente para promover la interacción de los estudiantes y lograr un aprendizaje común, el tipo de cooperación entre los estudiantes y la negociación que se realiza para llegar a acuerdos comunes sobre ideas, conceptos o procedimientos matemáticos.

### Prueba de funcionamiento del POAEM

La prueba de funcionamiento del protocolo de observación consistió en la aplicación del instrumento en la versión posterior al jueceo con especialistas en educación, matemáticas y metodología de diseño de instrumentos, así como en un pilotaje con un grupo de sujetos con características similares a la población objetivo. Por las características del estudio no se requirió una muestra representativa, pero sí contar con el número y tipo de sujetos necesarios para realizar las pruebas estadísticas apropiadas. En este caso, se contó con la participación de 20 docentes de quinto grado de primaria a quienes se videograbó en dos ocasiones cada uno. En total se contó con 40 clases videograbadas que a su vez se seccionaron en 222 actividades matemáticas, ya que éstas se tomaron como unidad de medida para el instrumento.

A partir de la prueba de funcionamiento se realizaron análisis estadísticos que ayudaron a determinar algunos aspectos de validez y confiabilidad de la información obtenida con el protocolo de observación. A continuación, se reportan los análisis estadísticos realizados y los resultados obtenidos en cada caso.

### Análisis Factorial Exploratorio y Confirmatorio

Para identificar la estructura dimensional del instrumento se realizó un Análisis Factorial Exploratorio (AFE) utilizando una base de datos con 111 casos aleatorios a través del programa estadístico SPSS, esto con el propósito de contar con información suficiente para realizar un AFE y posteriormente un Análisis Factorial Confirmatorio (AFC) con la misma cantidad de datos, pero con casos distintos. La factibilidad del tamaño de la muestra se definió con la prueba Kaiser, Meyer y Olkin (KMO) y la prueba de esfericidad de Bartlett (Tabla 1). En ambos casos los valores obtenidos fueron adecuados para realizar el AFE, ya que la medida de adecuación KMO fue mayor a .50 y el valor de significancia en la prueba de esfericidad fue menor a .05.

**Tabla 1:** Prueba KMO y esfericidad de Bartlett

KMO		.816
PRUEBA DE ESFERICIDAD DE	CHI-CUAD	204.295
	GL	21
BARTLETT	SIG.	.000

Para la extracción, rotación y selección de factores se utilizó el programa *R Factor para análisis factoriales* mediante el paquete SPSS. El método de extracción fue factorización de ejes principales, con una rotación oblicua (Quartimin), ya que este tipo de rotaciones genera soluciones más confiables y permite simplificar y clarificar la estructura de los datos. Los resultados del modelo factorial mostraron que el comportamiento de las cargas excluye la dimensión claridad del primer factor (Tabla 2). Esto significa que la dimensión claridad podría formar parte de una categoría distinta a la que agrupa a las seis dimensiones restantes.

**Tabla 2:** Análisis Factorial Exploratorio

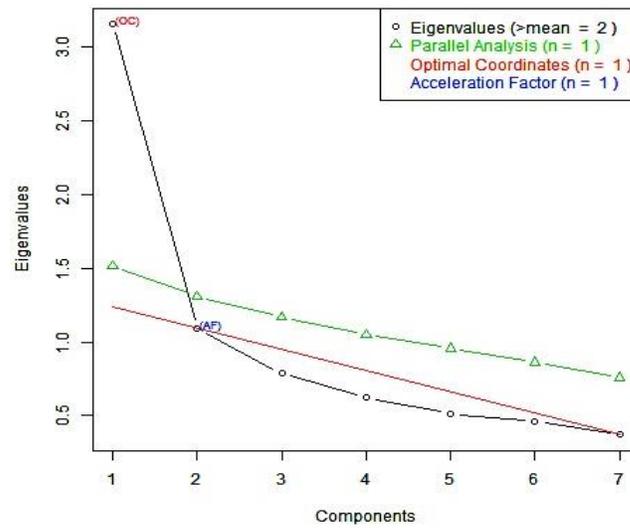
ÍTEM	(III CASOS)	
	F1	F2
CLARIDAD	.449	.682
EXIGENCIA	.746	-.034
RECURSOS	.658	.282
ALTERNATIVAS	.668	-.508
PREGUNTAS	.769	-.018
CONEXIÓN	.554	-.497
TRABAJO	.788	.205

El porcentaje de varianza explicada con el modelo de AFE excluyendo la dimensión “claridad de la tarea” fue del 50% (Tabla 3). No obstante, para determinar el número correcto de factores a retener con base en el AFE se utilizaron diferentes métodos: regla de Kaiser (autovalores), análisis de paralelo, coordinación óptima, factor de aceleración, y prueba de Velicer. Los resultados señalan que la regla de Kaiser retiene dos factores, mientras que el análisis paralelo establece el punto de corte en un factor, al igual que la coordinación óptima, la prueba Velicer, y el factor de aceleración (Figura 2).

**Tabla 3:** Porcentaje de varianza explicada y autovalores

FACTOR	AFE (III CASOS)		
	AUTOVALORES	% VARIANZA	ACUMULADA
1	3.009643	50.160711	50.160711
2	0.923707	15.395109	65.555820
GFI: .899			

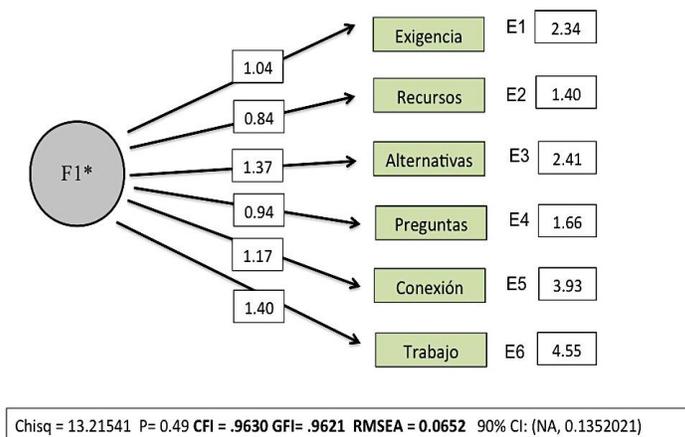
Figura 2: Estimación del número de factores a partir de cuatro métodos (todas las variables)



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

El modelo resultante del AFE se contrastó con un AFC utilizando el programa estadístico R y las librerías “Psych” y “Sem”. Los resultados mostraron que la estructura dimensional del modelo agrupa la mayoría de las dimensiones en un factor. En todos los casos se cumple con los supuestos para aceptar un modelo unifactorial, ya que los índices GFI= 0.96 y CFI= 0.96 presentan un buen ajuste, y el índice de error aproximado RMSEA=0.06 es poco menor al valor de referencia. Esto confirma la estructura identificada con el AFE en la Figura 3.

Figura 3: Modelo unifactorial.



Fuente: Diseño propio a partir de la estimación de parámetros.

### Consistencia interna: Alfa de Cronbach

Para calcular la correlación de cada reactivo o ítem con cada uno de los otros considerados en el POAEM se obtuvo el coeficiente de homogeneidad o consistencia interna Alfa de Cronbach. Los resultados sugieren que la consistencia interna es aceptable con un valor de 0.785 según los criterios de George y Melley (2003).

### Estudio de Generalizabilidad

Para ofrecer información sobre diferentes facetas de error en la medición con el POAEM, se realizó un estudio de confiabilidad basado en los componentes de varianza. El estudio de generalizabilidad O “estudio G” se basa en la confiabilidad del comportamiento de mediciones estimando múltiples fuentes de error de forma independiente. Los resultados ofrecen información sobre la confiabilidad para diferentes tipos de interpretación, ya que el universo de medición es definido por el conjunto de condiciones posibles en que se podría tomar la medida y al que se pretende generalizar el puntaje observado. Es decir, por la mayor fuente de variación a partir de las posibles fuentes de error en las muestras de una medida, por ejemplo: los ítems, juez, ocasión, segmentos, etcétera (Shavelson y Webb, 1991; Brennan, 2001).

Los datos recabados con el POAEM sugieren un diseño anidado desbalanceado con tres facetas, ya que de acuerdo con Shavelson y Webb “surge cuando el número de niveles de facetas anidadas varía en algún nivel, produciendo un número desigual de niveles” (2006, p. 610). En este caso, para el diseño  $P \times R \times O \times S:O$ , en donde “P” es el observado, “R” el juez, O la ocasión de observación y S:O el segmento dentro de ocasión, se tiene el mismo número de niveles para P, R y O; pero no para “S:O”.

La tabla 4 presenta los resultados para el conjunto de dimensiones que conforman el POAEM, en la primera columna se incluye la información del total de dimensiones del instrumento y en la segunda se excluye la dimensión “claridad de la tarea” tal como sugiere el AFE y el AFC. La información que se reporta incluye la proporción de varianza obtenida con el programa SPSS (ANOVA con efectos aleatorios), así como el porcentaje correspondiente a la varianza para cada una de las facetas del diseño  $P \times R \times O \times S:O$ , utilizando las fórmulas de cuadrados medios esperados E[MS].

Tabla 4: Varianza por dimensión para el POAEM

	TOTAL DE DIMENSIONES		DIMENSIONES SIN CLARIDAD DE LA TAREA	
	$\sigma^2$	%	$\sigma^2$	%
Docente (P)	23.085	29.5%	19.497	27.7%
Día (O)	-.725	-0.9%	-.627	-0.9%
Segmento (Día); (S:O)	.000	0.0%	.000	0.0%
Juez (R)	-.306	-0.4%	.107	0.2%
Docente * Día (PO)	-1.902	-2.4%	-.677	-1.0%
Docente * Segmento (Día); (PS:O)	17.170	21.9%	15.163	21.5%
Docente * Juez (PR)	24.258	31.0%	22.231	31.6%
Día * Juez (OR)	.763	1.0%	.529	0.8%
Juez * Segmento (Día); (RS:O)	-1.539	-2.0%	-1.432	-2.0%
Docente * Día * Juez (POR)	.347	0.4%	-.427	-0.6%
Docente * Juez * Segmento (Día), Error (PRS:O, e)	17.098	<b>21.9%</b>	16.002	<b>22.7%</b>
Total Varianza	78.249	100.0%	70.366	100.0%

Los resultados del conjunto de dimensiones muestran que el mayor porcentaje de varianza se concentra en las diferencias entre los jueces al calificar a los distintos docentes (PR) con 31%, seguido de la variación entre el desempeño de los docentes (P) con 30% y la variación de la práctica de los docentes entre segmentos (PS:O) con 22%. En total este diseño reporta una varianza residual (PRS:O,e) de 22%, por lo que el diseño con siete dimensiones explica el 78% de la varianza del modelo.

La misma situación se observa con el diseño  $P \times R \times O \times S:O$  excluyendo la dimensión “claridad de la tarea”, ya que los resultados reportan porcentajes menores a los obtenidos con todas las dimensiones. En este caso, el porcentaje de varianza más alto se concentra también en las diferencias entre los jueces al calificar a los distintos docentes (PR) con 32%, seguido de la variación entre el desempeño de los docentes (P) con 28% y la variación de la práctica de los docentes entre segmentos (PS:O) con 22%. El diseño con seis dimensiones reporta una varianza residual (PRS:O,e) de 23%, por lo que excluyendo la dimensión claridad es posible explicar 77% de la varianza total del modelo.

## Conclusiones

Las evidencias de validez de la información recabada con el POAEM se centraron en juicios de valor de un comité de expertos quienes discutieron elementos conceptuales y operacionales del instrumento, así como en un Análisis Factorial Exploratorio y un Análisis Factorial Confirmatorio para establecer la estructura dimensional del instrumento. Las evidencias de confiabilidad se establecieron con el coeficiente Alfa de Cronbach para determinar la consistencia interna, y con un estudio de generalizabilidad para identificar fuentes de error en la medición.

Las sesiones con los expertos permitieron asegurar que las dimensiones para medir el constructo estuvieran relacionadas con elementos conceptuales de las actividades de enseñanza. La estructura dimensional del instrumento se agrupó en un factor y no en cuatro como se había anticipado, además, la estructura factorial resultante excluyó la dimensión claridad de la tarea, ya que ésta no parecía medir aspectos relacionados con el constructo.

Con el estudio G se encontró que en conjunto la varianza de las dimensiones del POAEM explican cerca del 80% de la varianza. No obstante, de forma individual las dimensiones reportan porcentajes altos de error. Es decir, de forma independiente las dimensiones explican poco sobre las posibles fuentes de error en la medición (55%-60%).

Es importante mencionar que el diseño de instrumentos es un proceso complejo y laborioso que requiere de varias etapas de ajuste, de suerte que en algún momento sea posible sustentar generalizaciones. Para el POAEM será necesario modificaciones que comprendan la revisión de los criterios incluidos en las rúbricas de desempeño y reforzar un nuevo estudio de generalizabilidad que considere un diseño equilibrado. No obstante, se espera que los resultados obtenidos con este estudio y su proceso de desarrollo sean un referente para el desarrollo de nuevos instrumentos que busquen evaluar las prácticas docentes, ya que los estudios metodológicos son poco frecuentes en nuestro país, especialmente aquellos que buscan medir las prácticas de enseñanza.

Finalmente, se señalarán dos aportes de este estudio: 1) dimensiones específicas para valorar las actividades de enseñanza identificadas a partir de la revisión de literatura, las cuales son coincidentes con las orientaciones del modelo educativo para la educación obligatoria vigente al momento del levantamiento de los datos, y 2) la documentación del proceso de diseño y obtención de evidencias de validez y confiabilidad de la información recabada, ya que esto será un referente para otros estudios metodológicos que tengan como propósito diseñar instrumentos para evaluar las prácticas docentes.

## Referencias

- Aubuson, P., Burke, P., Schuck, S. y Kearney, M. (2014). Teachers choosing rich tasks: The moderating impact of technology on student learning, enjoyment, and preparation. En *Educational Researcher*, Vol. 43, No. 5, pp. 219-229
- Avery, P. y Freeman, C. (1999). Authentic instruction and assessment. *Social Education*. Recuperado de [http://www.eeraonline.org/journal/files/2002/JRE\\_2002\\_04\\_Avery.pdf](http://www.eeraonline.org/journal/files/2002/JRE_2002_04_Avery.pdf)
- Ávila et. al. (2013). Una década de investigación educativa en conocimientos disciplinares en México. Matemáticas, ciencias naturales, lenguaje y lenguas extranjeras 2002-2011. México: ANUIES-COMIE
- Boston, M. y Wolf, M. (2006). Assessing academic rigor in mathematics instruction: The development of instructional quality assessment toolkit. Reporte técnico 672. Recuperado de <http://www.cse.ucla.edu/products/reports/r672.pdf>
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.
- Hiebert, J. y Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal* 30, no 2. pp. 393-425
- Llinares, S. (2008). Aprendizaje del estudiante para profesor de matemáticas y el papel de los nuevos instrumentos de comunicación. *III Encuentro de Programas de Formación Inicial de Profesores de Matemáticas*. Santa Fe de Bogotá, Colombia: Universidad Pedagógica Nacional.
- Martínez Rizo, F. y Chávez Y. (2015). La enseñanza de matemáticas y ciencias naturales en educación básica en México. Revisión de literatura. Universidad Autónoma de Aguascalientes.
- McTighe, J. y Wiggins, G. (2012). Understanding by design framework. Consultado en: [http://www.ascd.org/ASCD/pdf/siteASCD/publications/UbD\\_WhitePaper0312.pdf](http://www.ascd.org/ASCD/pdf/siteASCD/publications/UbD_WhitePaper0312.pdf)
- McTighe, J., Seif, E. y Wiggins, G. (2004). You can teach for meaning. En *Educational Leadership*. 62 (1), pp. 26-31.
- National Council of Teachers of Mathematics. (2015). De los principios a la acción para garantizar el éxito matemático para todos. México: Libros S.A.
- Newmann, F., Bryk, A. y Nagaoka, J. (2001). *Authentic Intellectual Work and Standardized Test: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.
- Newmann, F., Marks, H. y Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, 104 (4), pp. 280-312.
- Newmann, F.M., Lopez, G. y Bryk, A.S. (1998) *The quality of intellectual work in Chicago schools: A baseline report*. Consortium on Chicago School Research, Chicago.
- Nicaise, M., Gibney, T., y Crane, M. (2000). Toward an understanding of authentic learning: student perceptions of an authentic classroom. *Journal of Science Education and Technology*, 9(1), 79-94.
- Picaroni, B. y Loureiro, G. (2010). ¿Qué matemática se enseña en aulas de sexto año de Primaria en escuelas de Latinoamérica?. *Páginas de Educación*, Vol. 3 (1). Recuperado de: <https://revistas.ucu.edu.uy/index.php/paginasdeeducacion/article/view/657>
- Santos, M. (2003). Hacia una instrucción que promueva los procesos de pensamiento matemático. En E. Filloy (coord). *Matemática Educativa: Aspectos de la investigación actual*, pp. 314-332. FCE: México.
- Secretaría de Educación Pública. (2011). Plan de Estudios. Educación Básica. México: SEP
- Secretaría de Educación Pública. (2017). Aprendizajes clave para la educación integral. Plan y programas de estudio para la educación básica: SEP

Shavelson, R. J., y Webb, N. M. (1991). *Generalizability Theory, A Primer*. Thousand Oaks, California: SAGE Publications, Inc.

Stein, M., Grover, B. y Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical task used in reform classroom. *American Educational Research Journal*, 33 455-488.

Swan, M. (2015). Designing tasks and lesson that develop conceptual understanding, strategic competence and critical awareness. Center for Research in Mathematics Education:University of Nottingham.

Verschaffel, Greer y De Corte (2010). Mathematics learning. *International Handbook of Mathematics Education*, Vol. 3 pp. 401-405.

Wiggins, G. y McTighe, J. (2008). Put understanding first. *En Educational Leadership*. 65 (8), pp. 36-41.