

LA EVALUACIÓN EN MATEMÁTICAS: EL CASO DE LA PRUEBA ESCRITA

Martha Imelda Jarero Kumul, Eddie Aparicio Landa

Facultad de Matemáticas de la Universidad Autónoma de Yucatán

jarerok@uady.mx, alanda@uady.mx

(México)

Resumen. En el presente trabajo se expondrán los resultados de un estudio desarrollado en una facultad de ciencias (ciencias exactas) cuyo propósito consistió en determinar la eficacia de una prueba escrita como recurso de medición de aprendizajes específicos, y validar sus alcances predictivos respecto a la complejidad de contenidos y ejercicios matemáticos que la componen. Con esto se esperaba caracterizar procesos, mecanismos e indicadores asociados a la elaboración de una prueba escrita.

Palabras clave: evaluación, prueba escrita

Abstract. In the present work we will expose the results of a study developed in a faculty of sciences (exact sciences) with the intention of determining the efficiency of a written test as resource of measurement of specific learning, and of validating his predictive scopes with regard to the complexity of contents and mathematical exercises that compose it. With this we are expecting to characterize processes, mechanisms and indicators associated with the production of a written test.

Key words: assessment, written test

Presentación de la problemática

¿Es posible determinar un valor predictivo en las pruebas escritas (pruebas abiertas) elaboradas por profesores de matemáticas universitarios en tanto la confiabilidad y validez de la complejidad resolutive? ¿La percepción de estudiantes universitarios sobre la complejidad resolutive de una prueba escrita se corresponde con los porcentajes de aciertos y desaciertos en la misma? Estas son cuestiones que en el presente escrito se discuten a partir de los resultados obtenidos en un estudio realizado con profesores en una facultad de ciencias exactas en México.

El interés por indagar sobre las formas en que profesores e instituciones educativas conciben y ejercen la evaluación de procesos educativos se ha intensificado en los últimos años. Las diversas propuestas incorporan reflexiones sobre el significado de la evaluación, sus procesos y usos.

Para algunos autores como Chadwick y Rivera (1997), la evaluación de procesos educativos requiere de coherencia entre los tipos de instrumentos o mecanismos usados y los objetivos de instrucción. Para algunos otros como Wachtel (1998) la evaluación es un proceso que se ve influenciado por factores relacionados con la administración de la misma, las características de los cursos, los perfiles de profesores y estudiantes. Para Runco (2003) la evaluación es un proceso selectivo y crítico en el cual las ideas originales, creativas y potencialmente útiles deben reconocerse y preferirse por encima de aquellas repetitivas, irrelevantes e inapropiadas.

En Sawin (2005) se plantea que las técnicas empleadas en el método científico pueden auxiliar en la evaluación que realizan los profesores, pues como se indica en Saroyan y Amundsen (2001), una forma de superar las limitaciones de una evaluación sumativa que a menudo realizan los profesores, es el empleo creativo de métodos y recursos con un enfoque multidimensional.

En este orden de ideas estuvimos interesados en estudiar y aportar evidencia empírica sobre la problemática de la evaluación de aprendizajes en matemáticas. Específicamente en el papel predictivo que puede establecerse en una prueba escrita de preguntas abiertas de un grupo de profesores universitarios y la correlación entre las percepciones de los estudiantes sobre la complejidad resolutoria de una prueba y sus respectivos porcentajes de aciertos y desaciertos.

Sin duda es preciso valorar la calidad y eficacia de prácticas educativas en matemáticas, especialmente en aspectos relacionados con formas de evaluar procesos educativos y las pruebas escritas, pues visiblemente la actividad de un profesor concentra la selección, organización, comunicación y evaluación de un conjunto de conocimientos institucionales. Respecto a esto último, es usual que el profesor universitario de matemáticas elabore y aplique pruebas escritas para determinar los estados aptitudinales de sus educandos referente a conocimientos, habilidades y “actitudes” adquiridas. Tal tipo de pruebas son a menudo empleadas con doble intencionalidad (muchas veces ajena a una idea de evaluación en un sentido amplio), por un lado, son un referente para determinar quiénes habrán de acreditar o no un curso y por otro lado, la nota (calificación) alcanzada en el mismo.

Evaluación y reprobación escolar

La reprobación y rezago escolar es un problema vigente en el sistema educativo mexicano que se acentúa conforme los jóvenes avanzan (promueven) en sus estudios. Así, según estadísticas del Instituto Nacional de Estadística y Geografía (INEGI) en el ciclo escolar 2006 – 2007, el porcentaje de reprobados en la matrícula corresponde al 4.1% de los estudiantes de primaria, 17.6% de secundaria y al 37.7% de bachillerato. Si la referencia se hace exclusivamente a los cursos de matemáticas, la situación se agudiza pues se reporta hasta el 50% de reprobación en las asignaturas del área de fisicomatemática en el nivel medio superior.

La reprobación escolar obstruye el avance de los jóvenes en su formación académica – científica. Hecho que se ha estudiado desde diferentes enfoques, por ejemplo desde un marco psicopedagógico, Robles y Martínez (2007) plantean que la reprobación en matemáticas se traduce en un problema con múltiples repercusiones, en un primer momento se afecta directamente la trayectoria escolar de los jóvenes al prolongarse su estadía en la escuela, asunto que se traduce en más gastos para el gobierno. De igual manera, en un segundo

momento se afecta directamente el estado financiero de las familias y más aun, se sitúa a los estudiantes en una etapa de desajustes emocionales y comportamientos derivados de la reprobación que, en muchos casos incide en lo vocacional al punto que la tendencia en la selección de carreras donde se consideran menos difíciles se ve incrementada.

Aun cuando la tendencia en las escuelas para determinar si un estudiante acredita un curso o avanza al siguiente nivel educativo incorpora distintas formas de evaluación, implicando con ello que el profesor se apoye de instrumentos diversos, en Melchor y Melchor (2002) así como en Aparicio, Jarero y Ávila (2007), se reporta que la mayoría de los profesores universitarios utilizan la prueba escrita como principal instrumento de “evaluación” y calificación. Situación que se hace más evidente en el área de ciencias matemáticas. En este sentido y con la pretensión de caracterizar procesos, mecanismos e indicadores asociados a la elaboración de una prueba escrita, nos dimos a la tarea de evidenciar la eficacia de una prueba escrita como recurso de medición de aprendizajes específicos, y validar sus alcances predictivos respecto a la complejidad de contenidos y ejercicios matemáticos que la integran.

Método de estudio

En el trabajo desarrollado se entiende que cualquier tipo de prueba que tenga por fin último, medir resultados de aprendizajes, debe, en mayor o menor medida, reunir una serie de características que den cuenta de su nivel de calidad para cumplir eficientemente la función para la cual ha sido creada. En Salkind (1999) por ejemplo, se plantea que una de las características indispensables de una “buena” prueba, es la validez de la misma, es decir, la medición de lo que se pretende medir; tómesese como caso, la precisión con que mide la conducta especificada en el objetivo sometido a verificación.

Al momento de analizar la utilidad de las pruebas de aprovechamiento o pruebas que muestran un área de conocimiento en particular, debe tomarse en cuenta como característica central la validez de contenido, que se define como el grado en que una prueba mide una muestra representativa del contenido de una asignatura o curso y la conducta que la manifiesta. El conocimiento sobre la validez de contenido de una prueba se verifica a través de la relación que presenta con los objetivos, por ello se debe verificar que los instrumentos de evaluación sean adecuados en nivel de dificultad, extensión y profundidad a lo expresado en los objetivos. Otra característica de una prueba es la confiabilidad. Se dice que una prueba es confiable o consistente, cuando al ser aplicada en diversas oportunidades, produce resultados aproximadamente similares. Ello da cuenta de la exactitud o precisión con que un instrumento mide algo.

Así, el método de investigación seguido en el estudio consistió en diseñar y aplicar una prueba escrita (prueba experimental), haciéndose acompañar de un análisis a priori y uno a posteriori. El diseño de la prueba se basó en la identificación de contenidos conceptuales y procedimentales que los profesores participantes en el estudio incluían en sus pruebas. Con esta actividad se daba validez al diseño de la prueba, pues se incluían reactivos equivalentes a los empleados por ellos.

El grupo de investigadores analizó cada uno de los reactivos considerados en la prueba experimental con base en la teoría de los esquemas propuesta por Marshall (citado en Real, Olea, Ponsoda, Revuelta y Abad, 1999), atendiendo la identificación de componentes de dificultad que intervienen en la resolución de los reactivos de una prueba de matemáticas, enfatizando tanto aspectos referidos en la consigna como en los implicados en el proceso resolutivo. La idea era producir un índice de dificultad para cada uno de los reactivos de la prueba así como de la prueba misma. El índice de dificultad se determinó en el intervalo $[0, 1]$ que se obtuvo de dividir los indicadores identificados en la consigna y los identificados en el proceso resolutivo, entre el total de los indicadores del reactivo. El valor arrojado permite interpretar que un reactivo resulta muy difícil, mientras más cercano a uno se encuentre o muy fácil al acercarse a cero.

Presentamos como ejemplo el análisis de un reactivo (ver tabla 1) en el cual se identificaron ocho indicadores que se presentan ya sea en la consigna y/o en la resolución, para hacer un total de 16 indicadores. Tanto en la consigna como en la resolución, se asigna el valor uno o cero si se encuentra presente o no el indicador, y para el caso que se ejemplifica se identificaron siete indicadores, tres correspondientes a la consigna y cuatro en la resolución. Así, el índice de dificultad asignado a dicho reactivo fue $7/16=0.43$, de modo que se interpreta como un reactivo de regular complejidad resolutiva dado que se aproxima al valor 0.50 que correspondería al valor medio del intervalo.

Reactivo 6. Dada la siguiente relación en \mathbb{Z} , $R = \{(m, n) \in \mathbb{Z}^2 \mid mn \geq 0\}$

Determinése si R es una relación de equivalencia. En caso de serlo, hállese el conjunto cociente.

COMPONENTES SEGÚN MARSHALL	INDICADORES	EN LA CONSIGNA	EN LA RESOLUCIÓN
I. Conocimiento de los rasgos distintivos de un fenómeno o situación	1.Relación de Equivalencia	0	1
	2.Producto cartesiano	1	0
	3.Desigualdad en los enteros	0	0

II. Conocimiento de las condiciones que los delimitan y por tanto, la posibilidad de aplicar determinado procedimiento para resolverlos	4.Determinar, inconsistencia o contraejemplo	I	I
	5.Desigualdad en los enteros	0	0
III. Conocimientos relativos a la planificación de la solución de un problema de una categoría dada (estrategias de resolución)	6.Propiedades de relación de equivalencia	0	I
IV. Conocimientos relativos a los algoritmos y reglas de cálculo necesarios para ejecutar el proceso de solución.	7.Desigualdad en los enteros	0	0
	8.Relación de pertenencia	I	I

Tabla 1. Análisis del reactivo 6 de la prueba experimental

La prueba quedó determinada con un total de ocho reactivos (ver tabla 2) buscando un promedio de índice de complejidad de la prueba adecuado o intermedio, y para ello se incluyeron reactivos cuyo índice de complejidad variaron entre 0.37 y 0.67.

Reactivo	Tema	Tipo de reactivo	Estructura	Índice de complejidad
1	Tipo de relaciones	Conceptual	Opción múltiple	0.37
2	Tipo de funciones	Conceptual		0.57
3	Composición de funciones	Procedimental		0.50
4	Tipo de funciones compuestas	Conceptual		0.42
5	Concepto función	Procedimental	Abierta	0.46
6	Relación de equivalencia	Procedimental		0.43
7	Función biyectiva	Procedimental		0.65
8	Inversa de una función	Procedimental		0.67

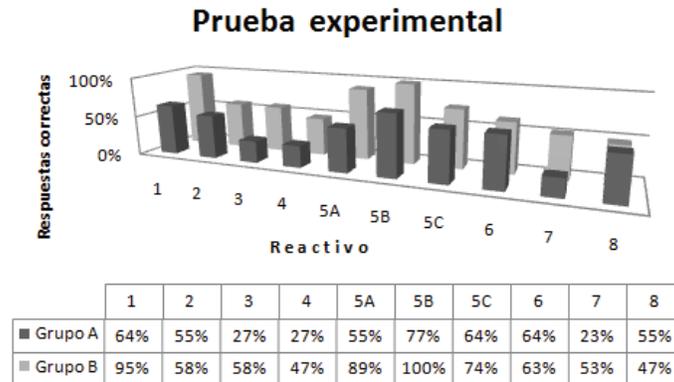
Tabla 2. Características de los reactivos de la prueba experimental

Resultados

La prueba fue implementada en una población de 41 estudiantes procedentes de dos grupos distintos, 22 estudiantes de un grupo etiquetado con la letra A y 19 de un grupo B. La implementación se realizó posterior a la presentación de los contenidos por parte de los profesores. Sin embargo cabe decir que el grupo B ya había pasado por el proceso de presentar la prueba que su profesor emplea para calificar el bloque de contenidos, situación distinta fue la del grupo A.

El tiempo asignado para responder la prueba experimental fue de 80 minutos. El grupo B tuvo un mejor desempeño al lograr porcentajes más altos en cada uno de los reactivos. Y aunque el

grupo A queda por debajo en el porcentaje de respuestas correctas, presenta un comportamiento similar al del grupo B, como puede verse en la gráfica 1.



Gráfica 1. Porcentaje de respuestas correctas por grupo

La diferencia encontrada se asoció al desfase de momento de aplicación de la prueba experimental respecto a la aplicación de la prueba del profesor.

Número de Reactivo	Índice de complejidad	Interpretación de los investigadores	Perspectiva del estudiante	% Respuestas Correctas
1	0.37	Fácil	Fácil a Muy fácil	78.0
2	0.57	Regular	Fácil a Muy fácil	56.1
3	0.50	Regular	Fácil a Regular	41.5
4	0.42	Regular	Regular a Fácil	36.6
5a	0.46	Regular	Regular a Fácil	70.7
5b	0.46	Regular	Regular a Fácil	87.8
5c	0.46	Regular	Regular a Fácil	68.3
6	0.43	Regular	Regular a Difícil	63.4
7	0.65	Difícil	Regular a Difícil	36.6
8	0.67	Difícil	Regular a Difícil	51.2

Tabla 3. Porcentaje de respuestas correctas de la prueba experimental

El poder predictivo de la prueba se analizó a partir de la interpretación de los estudiantes respecto al nivel de complejidad de cada uno de los reactivos, contrastado con lo propuesto por el grupo investigador así como los aciertos y desaciertos logrados por los estudiantes (tabla 3).

Los estudiantes consideraron que la prueba experimental incluía reactivos con diferente nivel de complejidad resolutoria, coincidiendo en la mayoría de los casos, con el índice de complejidad determinado por el grupo investigador. Sin embargo, aun cuando para los

estudiantes, los tres primeros reactivos resultaban perceptiblemente fáciles, éstos no resultaron ser los de mayor porcentaje de respuestas correctas.

Lo esperado era que el porcentaje de respuestas correctas sea inverso al índice de complejidad, esto es, que mientras más fácil esté catalogado un reactivo el porcentaje de respuestas correctas sea más alto. Tal situación no se cumple en lo general como puede verse de los resultados obtenidos en el reactivo 4. Este tiene menor porcentaje de respuestas correctas en tanto que el reactivo 6 que le sigue en dificultad tiene mayor porcentaje de respuestas correctas. De igual manera ocurre con los reactivos 5A y 5B, y entre el 7 y 8. Resultados que obligan a reconsiderar el índice de complejidad asignado y a la búsqueda de explicaciones más finas.

Conclusiones

De los resultados generados se concluye que aun cuando es posible generar una prueba escrita de preguntas y respuestas abiertas en matemáticas de educación superior confiable (en tanto se logra, exceptuando un reactivo, obtener resultados similares en poblaciones distintas), no es posible precisar un valor predictivo en la misma en su índice de complejidad resolutive.

Por otro lado se concluye que no fue posible bajo el método seguido, obtener una correlación entre la percepción de los estudiantes respecto a la complejidad resolutive de una prueba y sus porcentajes de logros. Lo que hace reflexionar sobre la conveniencia de ajustar el formato de la prueba, llevarla de una prueba abierta a una prueba cerrada de modo que el grado de logro se mida en términos dicotómicos: respuestas correctas y respuestas incorrectas, sin pasar por la consideración estados intermedios por parte del profesorado.

Referencias bibliográficas

- Aparicio, E., Jarero, M. y Ávila, E. (2007). La reprobación y rezago en cálculo. Un estudio sobre factores institucionales. *Premisa. Revista de la sociedad Argentina de Educación Matemática*. Edición 35, 3 – 12.
- Chadwick, C.B. y Rivera, N. (1997). ¿Qué es evaluación? En Chadwick, C. B. y Rivera, N. (Eds.). *Evaluación formativa para el docente* (pp. 36-61). México: Paidós Educador.
- Melchor, J. y Melchor, V. (2002). *El conocimiento de las matemáticas*. Recuperado el 30 de agosto de 2008 de <http://www.uaq.mx/matematicas/redm/art/a0904.pdf>
- Robles, R.; Martínez, J. (2007). La reprobación en matemáticas desde la perspectiva del alumno, el docente y la academia. *IPyE: Psicología y Educación* 1(1), 97-104

- Real, E.; Olea, J.; Ponsoda, V.; Revuelta, R. y Abad, F. (1999). Análisis de la dificultad de un test de matemáticas mediante un modelo componencial. Recuperado de <http://www.uv.es/revispsi/articulos2.99/4olea.pdf>
- Runco, M. A. (2003). Idea evaluation, divergent thinking, and creativity. To appear in M. A. Runco (Ed.), *Critical creative processes*. (pp. 69-94) Cresskill, NJ: Hampton
- Salkind, N. (1999). *Métodos de investigación*. México: Prentice Hall.
- Sawin, E (2005). The scientific method and other bases for evaluation procedures. *ETC*: 386 – 404.
- Saroyan, A. & Amundsen, C. (2001) Evaluating university teaching: time to take stock, *Assessment & Evaluation in Higher Education*, 26(4), 341–353.
- Wachtel, H. K. (1998) Student evaluation of college teaching effectiveness: a brief review, *Assessment & Evaluation in Higher Education*, 23(2), 191–211.