

CONSTRUCCIÓN Y VALIDACIÓN DE UNA PRUEBA PARA MEDIR CONOCIMIENTOS MATEMÁTICOS¹

Patricia López Obando²

Fecha de recepción: Junio 2009

Fecha de aceptación: Agosto 2009

Resumen

Se muestran los resultados de la investigación para la construcción y validación de una prueba previamente diseñada en pro de medir conocimientos en el área de matemática, desde una perspectiva clásica de la medición se llevo a cabo el proceso de validación por jueces, análisis de ítems y consistencia interna de dicha prueba, lo anterior por considerar la evaluación dentro del aula en la formación básica como un proceso de relevancia en un contexto social donde permanentemente se busca seleccionar, jerarquizar y diagnosticar mediante pruebas objetivas. Posterior al pilotaje llevado a cabo en un colegio privado en la ciudad de Bogotá con 131 estudiantes se encontraron falencias en las características técnicas de los ítems y por ende en el nivel de confiabilidad de la prueba, ello da muestra de la poca formación de los docentes para este tipo de procedimiento llevando a preguntarse respecto de la fiabilidad de la evaluación realizada.

Palabras Clave: evaluación, prueba objetiva, conocimiento matemático.

CONSTRUCTION AND VALIDATION OF A TEST TO MEASURE MATHEMATICAL KNOWLEDGE

Abstract

This article shows the research findings for the construction and validation of a test previously designed in behalf of measuring knowledge in the area of mathematics, from a classical perspective of the measurement himself I accomplish the process of validation for judges, analysis of items and internal consistency of happiness try, the above to regard the evaluation within the classroom in basic training as a process worthy of consideration in a social context where permanently it is tried to select, to hierarchies and to diagnose by means of objective tests. Posterior they found to the testing accomplished at a private school at the city from Bogotá with 131 students fallacies in the characteristic techniques of items and as a consequence in the level of reliability of the test, it gives the few sign formation of teachers for this procedural type taking to ask oneself in respect of the reliability of the realized evaluation.

Key words: evaluation, objective test, mathematical knowledge

INTRODUCCIÓN

La evaluación es una de las acciones que constantemente se realiza en la institución escolar, los resultados obtenidos permiten tomar decisiones con base en dos elementos: 1. determinar los aspectos que hay que mejorar y 2. definir y orientar las acciones que permitan elevar la calidad de la propuesta educativa (Instituto Colombiano Para el Fomento de la Educación Superior – ICFES, 2001). El evaluar por competencias supone la recolección de evidencia que permita determinar el saber hacer

de un sujeto en un contexto determinado, utilizando herramientas conceptuales y procedimentales adquiridas a través del proceso de formación.

Al evaluar por competencias se deben revisar los desempeños de los estudiantes, es decir, es una medición indirecta a partir de las actuaciones determinadas y esperadas de acuerdo al nivel manejado. Estos desempeños están directamente relacionados con el área a evaluar, se evidencian las características en el manejo de la información y su aplicación en situaciones problema definidas en el área de conocimiento.

¹ Investigación realizada en la especialización de Medición y Evaluación en Ciencias Sociales.

² Mg. En investigación Educativa. Especialista en Medición y Evaluación en Ciencias Sociales. Licenciada en Física. Coordinadora de Desarrollo Científico. Cooperación Universitaria Iberoamericana.

Según Rico (citado por González 2000, p. 21) "La evaluación es el enjuiciamiento sistemático de la valía o el mérito de un objeto. Esta definición se centra en el término valor e implica que la evaluación siempre supone un juicio". Lo anterior pone de manifiesto la relevancia de la evaluación en un contexto social ya que al ser un juicio permite seleccionar, jerarquizar, diagnosticar y medir procesos.

Es tal el impacto social que incluso la evaluación de las competencias básicas en las áreas de lenguaje y matemáticas ocupa un lugar prominente en el plan de desarrollo del gobierno colombiano, ya que se reconoce la necesidad inaplazable que tienen los ciudadanos de informarse de sí mismos sobre las debilidades y fortalezas en estas dos áreas (Barón, Rojas, Salazar, 2003).

Lo anterior surge del reconocimiento de la matemática como un ente globalizador de pensamiento lógico formal, dicho nivel de pensamiento permite evidenciar destrezas y habilidades surgidas en diferentes culturas como respuesta a diferentes necesidades sociales como contar, medir, localizar e inclusive jugar, lo que hace ver su practicidad no solo en un contexto académico sino también en un contexto cotidiano.

Por lo tanto al realizar una evaluación por competencias se busca trascender la perspectiva que la considera como la simple memorización de definiciones y ejecución de procedimientos o dominios de cálculo para llegar a valorar la destreza en el uso de este conocimiento en procesos de interpretación, representación y explicación de su entorno social (Barón et al, 2003a).

Se brindan a partir de lo anteriormente expuesto los elementos a tener en cuenta en la construcción de una prueba tendiente a medir el conocimiento dentro del área, la cual debe contener elementos de: comprensión de conceptos, formas de representación y uso de lenguaje matemático, considerados tanto en su componente conceptual como procedimental, utilizando para ello la formulación y resolución de situaciones problemas alejadas de una base memorística y de desarrollo de algorit-

mos básicos, para llegar a un verdadero análisis previo a la aplicación de los mismos. Es imposible desconectar los componentes conceptual y procedimental ya que estos se alimentan mutuamente, por ejemplo, se reconoce la potencia de las estructuras conceptuales para seleccionar un procedimiento adecuado en una situación específica con cierto nivel de complejidad (Barón et al, 2001).

Al hablar de evaluación en la educación básica secundaria al interior del aula se deben operacionalizar los componentes tanto conceptuales como procedimentales de manera tal que se puedan valorar las fortalezas y debilidades de los estudiantes en el área específica, en este caso en el área de matemáticas, exponiendo al sujeto a una situación problema que permita al docente determinar el grado de profundidad alcanzado con respecto a un criterio previamente establecido (Barón et al, 2003b).

Dichas situaciones problema pueden aplicarse tanto de forma grupal como individual, y pueden variar en su forma, desde talleres, laboratorios, revisiones bibliográficas, construcción de proyectos hasta evaluaciones de tipo objetivo que pueden ser orales o escritas de lápiz y papel, siendo estas últimas el interés principal de este estudio.

Según Herrera (2003), las pruebas de carácter objetivo son tradicionalmente instrumentos usados en la evaluación de conocimientos y habilidades. Generalmente están constituidas por preguntas cerradas y cumplen con las siguientes características:

1. La calificación no depende del evaluador, dado que dichas condiciones de asignación de puntajes están previamente estandarizadas,
2. Existe un control máximo de las condiciones de aplicación, que reduce la influencia de variables extrañas que pudiesen contaminar los resultados,
3. El proceso de diseño y construcción de la prueba contempla una serie de pasos organizados de manera sistemática y que implica trabajo interdisciplinario, tanto de expertos conceptuales como de expertos en técnicas de medición.

A este respecto cabe mencionar que en el contexto colombiano se han implementado evaluaciones de este tipo de forma masiva, ejemplo de ellos son las pruebas SABER aplicadas por la Secretaria de Educación Nacional y los exámenes de estado ICFES, en asocio con el Ministerio de Educación Nacional (MEN) y las instituciones de escolares del país.

Bajo este panorama se hace necesaria la proyección de este tipo de procedimientos de evaluación al interior del aula, buscando hacer un seguimiento constante de los estudiantes y potencializando de esta forma, las competencias pertinentes para su desempeño tanto en las pruebas masivas, como en situaciones de carácter evaluativo de su cotidianidad.

En el ambiente educativo, específicamente el nivel de educación básica formal, los docentes de forma un tanto empírica, construyen sus propias pruebas de evaluación, que son aplicadas, pero nunca son sometidas a procesos de validación y estandarización que permitan hacer inferencias de carácter sólido y científico; desde el mismo proceso de construcción se carece del apoyo de elementos de carácter teórico y técnico proporcionados por la teoría de la medición y evaluación, resultado de la carencia en formación específica en esta área, dentro de los programas de licenciatura y formación en pedagogía.

De acuerdo a la panorámica anteriormente presentada se hace necesario llevar a cabo procesos de construcción, validación y estandarización de las pruebas aplicadas en el aula de clase, siguiendo rigurosamente los lineamientos tradicionales de la medición y evaluación.

El aporte del presente estudio es entonces llevar a cabo la validación de una prueba en el área de matemáticas, por dos razones fundamentales: primero por considerar esta área como fundamental acorde con lo expuesto anteriormente y segundo pretende servir de modelo metodológico en procesos similares desde diversas áreas de la educación básica formal.

Dentro del proceso de construcción de una prueba

se contemplan las fases de: conceptualización, construcción de reactivos, validación, análisis de confiabilidad y por último la estandarización.

La fase de conceptualización inicia estableciendo la finalidad de la prueba en donde se especifica el constructo a medir la población y los objetivos del instrumento, por medio de la definición conceptual y operacional.

La fase de construcción de reactivos incluye el plan de prueba en donde se construye la tabla de especificaciones y se elige el formato de la prueba, es decir, tipo de ítems y estructura.

Al final de esta fase se lleva a cabo el análisis de ítems en donde se observan parámetros como índice de dificultad e índice de discriminación, retomando los parámetros establecidos para la aceptación de ítems planteados por Guilford (1975), donde el rango aceptable para el índice de dificultad oscila entre 0.2 y 0.85. Y en el caso de la discriminación retomando al mismo autor se asume un índice mínimo de 0.3

Y también se puede hacer un análisis de flujo de opciones con el fin de identificar la calidad de los distractores de dichos ítems. Retomando a Haladyna (1994) y Downing (1993) (citados por Rodríguez, et al, 2005). Se tendrá en cuenta el criterio de porcentaje de respuesta:

Poco razonables: si son respondidos por menos del 5% de los sujetos.

No discriminan: si son elegidos en igual proporción por quienes tienen puntajes altos y por quienes tienen puntajes bajos.

No claros: si varían a lo largo de los distintos rangos de puntajes de la prueba.

Indeseables: si imitan el comportamiento de la clave, es decir aumentan su frecuencia de respuesta a medida que aumentan los puntajes de la prueba. La fase de validación contempla la recolección de diferentes tipos de evidencia acorde con la finalidad

de la prueba, se encuentra aquí evidencia de validez por jueces expertos, análisis factorial, evidencia convergente, divergente, entre otros.

El análisis de confiabilidad se puede realizar desde una óptica interna o externa de acuerdo con las posibilidades de la investigación. En caso de poder hacer procedimiento con doble aplicación se puede utilizar el método de test -retest o pruebas paralelas, pero en caso contrario se lleva a cabo un proceso de consistencia interna utilizando división por mitades o coeficientes alfa, KR-20 entre otros.

El grado de integración de un ítem se puede descubrir investigando la relación que existe entre las respuestas que los sujetos emiten para el total de ítems. Lo que se investiga es la relación que existe entre el valor de escala que nos proporcionaría un ítem dado y el valor de escala que nos proporcionarían todos los ítems (Cohen y Swerdilk, 2001). Una relación baja o negativa será una indicación de que el ítem no está bien integrado, por el contrario una relación moderada o alta es una indicación de que la integración del ítem es aceptable o excelente.

Posterior al análisis ítem-test es relevante realizar el análisis ítem-ítem para verificar la relación existente entre los reactivos que componen el instrumento de medición (Cohen y Swerdilk, 2001).

El procedimiento es relevante ya que permite una estimación inicial del grado de consistencia que puede tener una prueba dado que entre mayor sean los índices de correlación entre los reactivos, mayor será la probabilidad de tener un instrumento con una adecuada consistencia o estructura interna (Cohen y Swerdilk, 2001).

El procedimiento siempre debe realizarse después del análisis ítem-test, es decir, se hace después de la depuración de los reactivos iniciales del instrumento.

En la fase de estandarización se determinan las normas generales tanto de aplicación como de calificación e interpretación de la información obtenida en el instrumento o prueba elaborada.

MÉTODO

Se utilizó el instrumento construido en el colegio Agustiniانو Ciudad Salitre (C.A.C.S) de la ciudad de Bogotá. Dicho instrumento está dirigido a estudiantes de grado décimo de educación básica formal y tiene como objetivo general evaluar el nivel de conocimiento en trigonometría, acorde con el plan de estudios del nivel. Es necesario aclarar que este estudio retomó el instrumento ya construido por los docentes a cargo de esta área en dicha institución educativa.

Para la fase de validación por jueces se contó con la participación de 10 jueces expertos en el área de matemáticas, quienes contaban mínimo con título de pre-grado en Licenciatura en matemáticas.

Para la fase de pilotaje se contó con la participación de 131 estudiantes pertenecientes al grado décimo del C.A.C.S, 73 hombres y 58 mujeres con edades comprendidas entre los 14 y 17 años.

Para la fase de validación por jueces se utilizó un formato que pretendía evaluar la claridad y pertinencia de cada uno de los 20 ítems incluidos en la prueba inicial.

Para la fase de análisis de ítems y análisis de consistencia se aplicó la prueba piloto ajustada con base en los ajustes de los jueces expertos.

El presente estudio se llevó a cabo mediante las siguientes fases:

FASE DE CONSTRUCCIÓN: Pese a iniciar la investigación con una prueba ya construida por un grupo de docentes del área de matemáticas del colegio, se realizó el proceso de validación por jueces y análisis de ítems correspondientes a la fase de construcción.

VALIDACIÓN POR JUECES: Se organizaron cada uno de los 20 ítems originales de la prueba en un cuadernillo de validación por jueces, con el fin de determinar la pertinencia y la claridad de los ítems por medio de la RAZÓN DE VALIDEZ DE CON-

TENIDO (CVR) de cada uno de los ítems.

AJUSTES INICIALES: Con base en el nivel de acuerdo de los jueces se eliminaron los ítems con niveles de valores no permisibles de CVR, acorde con los criterios de Lawshe (1975), (citado por Cohen, 2001).

PILOTAJE: Se aplicó la prueba a los 131 estudiantes de grado décimo del C.A.C.S. previamente divididos en 4 grupos correspondientes a los cuatro cursos del nivel, Las condiciones de aplicación se igualaron en tiempos (90 minutos), espacios (salón de clases) y herramientas de ayuda (calculadora).

ANÁLISIS DE REACTIVOS: Basados en el análisis clásico se llevó a cabo el análisis ítem – ítem, índice de dificultad, índice de discriminación, ítem – prueba, de cada uno de los 17 ítems piloteados.

ANÁLISIS DE CONFIABILIDAD: con los datos obtenidos en el pilotaje y dado que no era posible llevar a cabo un método de doble aplicación, se decidió realizar un análisis de consistencia interna implementando los métodos de: división por mitades, KR-20, Rulkon y Guttman. Se utilizó el paquete estadístico SPSS 10.0

RESULTADOS

Análisis de Ítems

Las características evaluadas en los ítems son su dificultad y su discriminación.

Se retomaron los parámetros establecidos para la aceptación de ítems planteados por Guilford (1975), donde el rango aceptable para el índice de

TABLA 1. Descriptivos de los parámetros de dificultad y discriminación con métodos de la TCT, para cada uno de los 20 ítems

ITEM No.	DIFICULTAD	CORR. BISERIAL PUNTUAL
1	.409	.41
2	.63	.16
3	.23	.25
4	.36	.37
5	.77	.34
6	.16	.38
7	.77	.29
8	.37	.37
9	.81	.29
10	.86	.53
11	.91	.12
12	.63	.41
13	.68	.30
14	.83	.39
15	.44	.40
16	.60	.38
17	.51	.46
18	.40	.32
19	.28	.17
20	.14	.13

dificultad oscila entre 0.2 y 0.85. Y en el caso de la discriminación retomando al mismo autor se asume un índice mínimo de 0.3.

Acorde con los parámetros de Guilford (Citado por Rodríguez, Casa, Medina, 2005). los ítems 6, 10, 11 y 20 presentan índices de dificultades no

adecuadas y por no discriminar los ítems 2, 3, 7, 9, 11, 19 y 20.

Como resultado del análisis del flujo de alternativas se enuncian en la tabla 2, los ítems que se entraron a considerar por presentar un comportamiento no adecuado.

TABLA 2. Comportamiento de la clave y los distractores de los ítems de la prueba.

ITEM #	ANÁLISIS DE ALTERNATIVAS
1	La clave a es contestada por la mayoría, sin embargo la opción d se comporta en forma muy similar y además la opción c solamente es elegida por 2 sujetos
3	La clave no resulta atractiva para los sujetos, ya que solo es elegida por el 16% , mientras que el distractor c es elegido por el 18 % de los individuos
6	La clave presenta la frecuencia de respuesta más baja, mientras que el distractor a fue el que más eligieron los sujetos
7	La clave es demasiada obvia ya que es elegida por el 71% de la población
8	La clave presenta un comportamiento muy similar al distractor d
9	La clave es demasiada obvia ya que es elegida por el 81% de la población
10	El distractor a es demasiado obvio ya que no tuvo frecuencia de respuesta y la clave es demasiado obvia al ser elegida por el 87% de los sujetos.
11	El distractor c es demasiado obvio ya que no tuvo frecuencia de respuesta y la clave es demasiado obvia al ser elegida por el 89% de los sujetos.
12	El distractor d es demasiado obvio ya que solo un sujeto lo eligió, la clave presenta un comportamiento adecuado.
13	El distractor a es demasiado obvio ya que no tuvo frecuencia de respuesta.
14	La clave es demasiada obvia ya que es elegida por el 81% de la población
15	La clave a tiene un comportamiento muy similar al distractor b, mientras que los distractores c y d son muy obvios al presentar frecuencias de respuestas muy bajas.
16	El distractor b es demasiado obvio al ser contestado únicamente por 2 sujetos.
17	El distractor b es demasiado obvio al ser contestado únicamente por 1 sujeto.
19	El distractor a presenta mayor frecuencia de respuesta en comparación con la clave
20	El distractor d atrae mayor atención al ser elegido por el 33%, mientras que la clave es elegida por el 9% de los sujetos

Con respecto al análisis ítem test al utilizar el coeficiente de Pearson se obtuvieron los siguientes resultados:

TABLA 3. Índices de correlación Ítem- Prueba

ITEM	ÍTEM □ TOTAL CORRELATION	ÍTEM	ÍTEM □ TOTAL CORRELATION
1	.2654	11	.1348
2	.0874	12	.3097
3	.1689	13	.2596
4	.0837	14	.3827
5	.1655	15	.1959
6	.1664	16	.1643
7	.1741	17	.3227
8	.2069	18	.1695
9	.3567	19	.1664
10	.3867	20	-.0962

Lo anterior permite observar que de 20 ítems que comprenden la prueba piloteada solamente los ítems 9, 10, 12, 14 y 17 se relacionan adecuadamente con el puntaje total de la prueba y el resto de ítems no están bien integrados a la prueba, especialmente el ítem 20 que arroja una correlación negativa.

Posterior a este análisis ítem -test o ítem prueba se procede a realizar el análisis ítem-ítem con el fin de iniciar el proceso de recolección de evidencia que permita establecer como se relacionan los ítems entre si. Los resultados se muestran en la tabla 4.

TABLA 4.

	item1	Item2	item3	item4	item5	item6	item7	item8	item9	item10	item11	item12	item13	item14	item15	item16	item17	item18	item19	item20
item 1	1,000	0,278	0,039	-0,059	0,174	0,183	0,084	0,113	0,232	0,134	0,013	0,510	0,094	0,083	0,134	0,012	0,167	0,052	0,012	-0,164
item 2	0,278	1,000	0,235	-0,143	0,132	-0,072	0,068	-0,036	0,149	-0,092	0,031	-0,027	0,030	0,071	0,013	-0,010	0,045	0,049	-0,030	-0,176
item 3	0,039	0,235	1,000	-0,081	0,109	0,007	0,004	0,133	0,076	0,084	-0,086	0,032	0,123	0,138	0,208	-0,073	0,002	0,102	0,061	-0,103
item 4	-0,059	-0,143	-0,081	1,000	0,064	-0,025	0,146	-0,099	-0,057	0,137	0,024	0,065	-0,021	0,151	-0,006	0,043	0,096	0,122	0,065	0,238
item 5	0,174	0,132	0,109	0,064	1,000	0,053	0,175	0,114	0,208	0,155	0,004	0,120	0,076	0,124	-0,009	-0,001	-0,042	-0,119	-0,077	-0,259
item 6	0,183	-0,072	0,007	-0,025	0,053	1,000	-0,105	0,151	0,018	0,020	0,097	0,185	0,067	0,039	0,144	0,113	0,102	0,022	0,126	-0,097
item 7	0,084	0,068	0,004	0,146	0,175	-0,105	1,000	-0,035	0,168	0,132	-0,134	-0,003	0,305	0,116	0,049	-0,105	0,086	0,055	-0,006	0,019
item 8	0,113	-0,036	0,133	-0,099	0,114	0,151	-0,035	1,000	0,191	0,071	0,085	0,294	0,138	0,056	0,049	0,131	0,049	0,015	0,067	-0,184
item 9	0,232	0,149	0,076	-0,057	0,208	0,018	0,168	0,191	1,000	0,203	0,186	0,212	0,187	0,153	-0,056	0,066	0,235	0,084	0,113	0,130
item 10	0,134	-0,092	0,084	0,137	0,155	0,025	0,132	0,071	0,203	1,000	0,047	0,191	0,050	0,492	0,121	0,177	0,208	0,197	-0,036	0,153
item 11	0,013	0,031	-0,086	0,024	0,004	0,097	-0,134	0,085	0,186	0,047	1,000	0,112	0,043	0,016	0,238	0,162	0,140	-0,067	0,033	-0,172
item 12	0,051	-0,027	0,032	0,065	0,120	0,185	-0,003	0,294	0,212	0,191	0,112	1,000	0,093	0,249	-0,017	0,122	0,176	0,058	0,101	-0,014
item 13	0,094	0,030	0,123	-0,021	0,076	0,067	0,350	0,138	0,187	0,050	0,043	0,093	1,000	0,066	0,215	0,067	0,050	-0,025	0,067	0,008

	item1	Item2	item3	item4	item5	item6	item7	item8	item9	item10	item11	item12	item13	item14	item15	item16	item17	item18	item19	item20
item 14	0,083	0,071	0,138	0,151	0,124	0,039	0,116	0,056	0,153	0,492	0,016	0,249	0,066	1,000	0,066	0,061	0,168	0,316	0,039	0,088
item 15	0,134	0,013	0,208	-0,006	-0,009	0,144	0,049	0,049	-0,056	0,121	0,238	-0,017	0,215	0,066	1,000	0,144	0,133	-0,139	0,102	-0,041
item 16	0,012	-0,010	-0,073	0,043	-0,001	0,113	-0,105	0,131	0,066	0,177	0,162	0,122	0,067	0,061	0,144	1,000	0,113	-0,014	0,072	0,031
item 17	0,167	0,045	0,002	0,096	-0,042	0,102	0,086	0,049	0,235	0,208	0,140	0,176	0,050	0,168	0,133	0,113	1,000	0,186	0,186	0,008
item 18	0,052	0,049	0,102	0,122	-0,119	0,022	0,055	0,015	0,084	0,197	-0,067	0,058	-0,025	0,316	-0,139	-0,014	0,186	1,000	0,163	0,123
item 19	0,012	-0,030	0,061	0,065	-0,077	0,126	-0,006	0,067	0,113	-0,036	0,033	0,101	0,067	0,039	0,102	0,072	0,186	0,163	1,000	0,031
item 20	-0,164	-0,176	-0,103	0,238	-0,259	-0,097	0,019	-0,184	-0,130	0,153	-0,172	-0,015	0,008	0,088	-0,041	0,031	0,008	0,123	0,031	1,000

La tabla 4 permite observar las correlaciones entre los ítems que mantienen concordancia con los resultados anteriores, ya que son muy débiles, a excepción de los ítems 9, 10, 12, 14 y 17 se evidencia un comportamiento disgregado de los elementos que constituyen la prueba.

Continuando con el proceso de análisis de la prueba se presentan los coeficientes de consistencia interna utilizados como evidencia de consistencia y confiabilidad en la prueba. Se aplicó el método de división por mitades, alfa de Cronbach, coeficiente alfa estandarizado y coeficientes θ y Ω .

A partir de los diferentes métodos se obtuvieron valores de consistencia muy bajos para una prueba de conocimientos, siendo el menor 0.54 y el mayor 0.64.

DISCUSIÓN

Los resultados permiten confirmar la problemática expuesta respecto de las dificultades en el diseño y construcción de pruebas de conocimiento matemático validas al interior del aula de clase. Ya que como queda expuesto en la presente investigación las pruebas realizadas no permiten y no son usadas en la toma de decisiones, no determinan los aspectos a mejorar y no son usadas por los maestros para definir y orientar acciones tendientes a elevar la calidad de los procesos educativos, es decir, no cumplen con los criterios necesarios para ser considerada una evaluación válida. Lo anterior debido a la falta de capacitación al respecto por parte de los docentes.

En cuanto a la construcción de la prueba; en los análisis ítem-prueba e ítem - ítem se observan correlaciones muy bajas que indican la falta de relación de los ítems entre sí y de los ítems con la prueba. Puesto que las pruebas utilizadas no atienden al proceso y criterios para la construcción de la misma, a saber: conceptualización, construcción de reactivos, validación, análisis de confiabilidad y estandarización. La prueba presentó dificultades desde el mismo proceso de construcción de los ítems evidenciado esto en el comportamiento poco deseable de 18 de los 20 ítems que presentaban dificultades tanto en las clave como en los distractores. Se observaron niveles de dificultad que tendían a los extremos y un número importante de ítems que no discriminan. A este respecto es importante apuntar que se observó una gran cantidad de datos perdidos o preguntas sin responder lo que pudo afectar la representatividad de estos dos parámetros.

En cuanto a los resultados obtenidos, se demuestra el olvido o poca importancia que se da a la observación y medición de las competencias matemáticas; las respuestas obtenidas no dan referencia del nivel de análisis, razonamiento o uso de los conceptos en la resolución de problemas matemáticos, de otra parte, para el estudiante es poco importante el rol de la matemática en su hacer cotidiano, la prueba se convierte así en un simple examen a responder.

En cuanto al análisis de confiabilidad, se eligió utilizar el análisis de consistencia interna debido a las dificultades para llevar a cabo una doble aplicación

de la prueba; por el método de división de mitades se obtuvo un valor de 0.54, el alfa de Cronbach arrojó un valor de 0.06037, el alfa estandarizado es de 0.62, Coeficiente θ es de 0.60 y el coeficiente Ω un valor de 0.64; lo que permite deducir que la prueba no es consistente, es decir, no mide de forma correcta el rasgo para el cual se construyó, además de no controlar la influencia del error implícita en todo proceso de medición.

En el momento de dar inicio al trabajo de validación no se encontró definición conceptual y operacional del constructo a medir, así como tampoco una tabla de especificaciones que permitiera asegurar la calidad en el proceso de construcción de los ítems y de la prueba en sí, por lo tanto los resultados encontrados ratifican que esta prueba no cuenta con los requerimientos técnicos exigidos por la medición para poder ser utilizada en un ambiente educativo con fines de evaluación, lo que pone de manifiesto dos situaciones:

1. Se están desarrollando procesos de evaluación erróneos al interior del aula; en tanto no dan cuenta del proceso y adquisición de competencias y no se convierten en criterios para la toma de decisiones, restringiéndose a la "medición" de conceptos gregarios y fuente de simples notificaciones para maestro, alumno y padres de familia, y
2. Es necesaria la formación docente respecto de la elaboración de pruebas validas para evaluar conocimiento; la gran mayoría de maestros no cuenta con la preparación conceptual como para asumir la evaluación en todo su proceso, o por otra parte teniendo la preparación para ello, no cuenta con el tiempo que permita la rigurosidad para realizarla. Finalmente responde a las necesidades del momento; el registro de notas y los procesos institucionales.

Lo anterior podría ser objeto de una nueva investigación, no menos exhaustiva e interesante

pues ¿cuáles son las razones que llevan a un maestro consciente de los criterios y procesos necesarios para realizar un proceso real de evaluación a obviarlos y no realizarla? Número de estudiantes a cargo, tiempo de preparación y ejecución, cabe anotar aquí, que la presente investigación aunque esta realizada en el área de matemáticas, da cuenta de las dificultades en todas las áreas del conocimiento y podría aplicarse igualmente en todas ellas. De hecho en la institución educativa aplicada recibió la atención de todos los maestros, quienes expresaban la necesidad de iniciar un proceso tanto de diagnóstico como de aplicación y finalmente de preparación de los docentes en la realización de pruebas válidas.

REFERENCIAS

- Barón C., Rojas P., Salazar C. (2003a). Matemáticas escolares: aportes para orientar procesos de innovación. Bogotá: ICFES Instituto Colombiano Para el Fomento de la Educación Superior ICFES.
- Barón C., Rojas P., Salazar C. (2003b). Evaluar para Transformar: Una mirada a los fundamentos e instrumentos de matemáticas 2002 - 2003. Bogotá: ICFES Instituto Colombiano para el Fomento de la Educación Superior ICFES.
- Cohen, R., Swerdilk, M. (2001). Pruebas y Evaluación Psicológica. Introducción a las pruebas y a la Medición. Mexico: Mc Graw Hill.
- González S., (2000). Acerca de la evaluación en matemáticas. En revista Educación y Cultura. Instituto Colombiano para el Fomento de la Educación Superior (2003) Evaluar para transformar -aportes de las pruebas saber al trabajo en el aula- Colombia. Bogotá: ICFES.
- Instituto Colombiano para el Fomento de la Educación Superior (2006). Prueba de matemática. Recuperado en <http://www.icfes.gov.co>. Colombia.
- Rodríguez, O., Casa, P., Medina, Y. (2005). Análisis psicométrico de los exámenes de evaluación de la calidad de la educación superior (Ecaes) en Colombia. *Revista Avances en Psicología*. Volumen 11. N° 1. 2009 / págs. 29-37

