

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Educación



Estudio de tres tipos de validez de una prueba de matemáticas de tercer grado primaria

Trabajo de graduación en modalidad de tesis presentado por

Karla Magaly Roldán de la Rosa de Rojas

para optar al grado académico de Maestría en
Medición, Evaluación e Investigación Educativa

Guatemala

2015

Estudio de tres tipos de validez de una prueba de matemáticas de tercer grado primaria

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Educación



Estudio de tres tipos de validez de una prueba de matemáticas de tercer grado primaria

Trabajo de graduación en modalidad de tesis presentado por

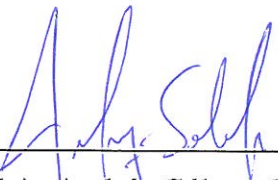
Karla Magaly Roldán de la Rosa de Rojas

para optar al grado académico de Maestría en
Medición, Evaluación e Investigación Educativa

Guatemala

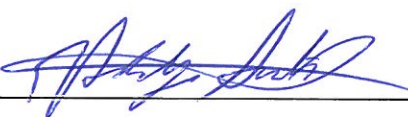
2015

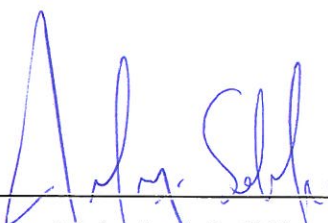
Vo. Bo.

(f) 
M.A. Andrés Gálvez-Sobral (Asesor)

Tribunal Examinador:

(f) 
M.A. Bayardo Mejía Monzón

(f) 
M.A. Adolfo Santos

(f) 
M.A. Andrés Gálvez-Sobral

Fecha de aprobación: Guatemala, 28 de mayo de 2015 ✓

PREFACIO

El presente trabajo de investigación surge a partir del curso de trabajo de campo realizado por la autora del mismo sobre un estudio de validez del nivel cognitivo de ítems de una prueba de matemáticas de tercer grado, que inicialmente fue planteado bajo un enfoque cualitativo. Por interés personal y como sugerencia del asesor del mismo y director del Centro de Investigaciones Educativas de la Universidad del Valle de Guatemala, se buscó agregar un enfoque cuantitativo añadiendo la parte de validez predictiva y de constructo de la misma prueba, con lo cual la investigación quedó conformada con un carácter mixto.

El Centro de Investigaciones Educativas (CIE) de la Universidad del Valle de Guatemala ha desarrollado pruebas de lectura y matemáticas de tercer grado, las cuales han sido sometidas a diferentes procesos de validación interna y con distintas muestras de sujetos. A partir de ello, surge la necesidad de contar con información que refleje las fuentes de validez de criterio, de constructo y de contenido (del nivel cognitivo) de la prueba de matemáticas de tercer grado, debido a que para el CIE es pertinente contar con estudios de validez que aumenten el rigor científico de las pruebas que se utilizan.

Por tal motivo se agradece la información proporcionada por el CIE, en relación a las bases de datos de resultados de evaluaciones aplicadas a niños de tercero y sexto grado de primaria, para la realización de la presente investigación. Asimismo, se reconoce el aporte brindado por el grupo de profesionales en educación que fungieron como expertos en la validación del nivel cognitivo de las respuestas de los niños, quienes proporcionaron la información por medio de audios y transcripciones de los mismos. Ellas son Ana Lucía Asturias, Karla Archila, Lorena Bianchi, Anaidé Cruz y Silvia Rosal.

CONTENIDO

	Página
PREFACIO	v
LISTA DE TABLAS	vii
LISTA DE GRÁFICOS	viii
RESUMEN	ix
I. INTRODUCCIÓN	1
II. DEFINICIÓN DEL PROBLEMA	2
III. MARCO TEÓRICO	3
IV. MARCO METODOLÓGICO	34
V. RESULTADOS	39
VI. DISCUSIÓN	61
VII. CONCLUSIONES	66
VIII. RECOMENDACIONES	72
IX. BIBLIOGRAFÍA	73
X. APÉNDICES	77

LISTA DE TABLAS

Tabla	Página
1. Estadísticos descriptivos	40
2. Correlaciones	40
3. Regresión	41
4. Comunalidades	44
5. Varianza total explicada	46
6. Matriz de componentes rotados	50
7. Factores	54
8. Ítems alineados	56
9. Ítems parcialmente alineados	58
10. Ítems no alineados	59
11. Ítems sin evidencia de alineación	60
12. Reclasificación de ítems	71

LISTA DE GRÁFICOS

Gráfico	Página
1. Gráfico de correlación	41
2. Gráfico de sedimentación	48

RESUMEN

De acuerdo a la última edición de los Estándares del comité conjunto de la Asociación Americana de Psicología, la Asociación Americana de Investigación Educativa y el Consejo Nacional de Mediciones en Educación (edición de 1999 de la APA, AERA y NCME) citada en Padilla, Gómez, Hidalgo y Muñiz (2006), la validez se refiere a “un juicio integrado del grado en el que la evidencia empírica y la racionalidad teórica apoyan la suficiencia y conveniencia de las inferencias y acciones basadas en las puntuaciones en los tests u otros modos de evaluación”. En ese sentido, en este estudio se pretende proporcionar fuentes de evidencia de la validez de los resultados obtenidos en la aplicación de la prueba de matemáticas de tercer grado que utiliza el Centro de Investigaciones Educativas –CIE- de la Universidad del Valle de Guatemala, mediante los análisis de validez predictiva y de constructo, así como el método de validez cognitiva “pensar en voz alta” (think aloud) al razonar los ítems de la prueba, a través del uso de protocolos verbales o evidencia sobre los métodos que utilizan las personas cuando toman una prueba, al momento de resolver las tareas que ello implica (Thelk, Hoole y Lottridge, s.f.).

Se realizó el análisis de validez predictiva mediante la correlación de los resultados obtenidos en la prueba de matemáticas de tercer grado en 2011 y los resultados obtenidos por los mismos niños ($n=126$) en la prueba de matemáticas de sexto grado en 2014. Se encontró una correlación con un valor de 0.646 ($p<0.01$) y un valor R^2 de 0.417, es decir los resultados de la prueba de matemática de sexto grado dependen en un 41.7% de los resultados de la prueba de matemática de tercer grado.

El análisis de validez de constructo se realizó con la muestra total de niños de tercer grado ($n=237$) evaluados en el 2011, por medio de un análisis factorial confirmatorio, a través del programa SPSS versión 20, con el método de extracción de componentes principales sobre cuatro factores, como inicialmente había sido planteado en la tabla de especificaciones de la prueba, de acuerdo a la taxonomía de Marzano, en donde el porcentaje de la varianza total explicada es de 45.6%, de esta forma: 15.98% en el factor 1

Conocimiento, 11.62% en el factor 2 Comprensión, 9.92% en el factor 3 Utilización y 8.08% en el factor 4 Análisis.

El análisis de validez del nivel cognitivo de los ítems de la prueba, se llevó a cabo a través de las respuestas orales de los niños (n=21) en el año 2012, que fueron grabadas y posteriormente transcritas para el respectivo análisis de contenido, en donde se encontró un 60% de “alineación” total de la prueba entre lo planteado por los creadores de la misma según la tabla de especificaciones, los expertos en contenido (conformados por cinco profesionales del área de educación estudiantes de maestrías en educación de la Universidad del Valle de Guatemala) y las respuestas orales de los niños.

Finalmente se incluye la recomendación de reclasificar el nivel cognitivo de los ítems 6, 8, 14, 18, 28 y 29 de la prueba de matemática de tercer grado, debido a que la evidencia obtenida en el análisis factorial, así como el resultado de alineación cognitiva de los ítems, no coincide con el nivel cognitivo propuesto en la tabla de especificaciones de la prueba elaborada por los creadores de la misma.

I. INTRODUCCIÓN

La investigación se presenta como trabajo de graduación en modalidad de tesis, en la cual se realiza el análisis de validez predictiva y de constructo de una prueba de matemáticas de tercer grado que utiliza el Centro de Investigaciones Educativas de la Universidad del Valle de Guatemala. Se hace referencia a la taxonomía de Marzano y los diferentes dominios, niveles y procesos del sistema de cognición del ser humano. Asimismo se enfatiza en el método de validez cognitiva “pensar en voz alta” (think aloud) al razonar ítems de la prueba; todo ello debido a que para el CIE es pertinente contar con estudios de validez que aumenten el rigor científico de las pruebas que utilizan.

El análisis de validez predictiva se lleva a cabo mediante la correlación de los resultados obtenidos en la prueba de matemáticas de tercer grado y los resultados obtenidos por los mismos niños en la prueba de matemáticas de sexto grado. El análisis de validez de constructo se realiza por medio de un análisis factorial confirmatorio con el método de extracción de componentes principales de los datos obtenidos en la prueba. Por último, el análisis de validez del nivel cognitivo de los ítems de la prueba, se lleva a cabo a través de las respuestas orales de los niños que fueron grabadas y posteriormente transcritas para el respectivo análisis de contenido.

El aporte principal de la investigación es proporcionar fuentes de evidencia de validez de la prueba de matemáticas de tercer grado, para ampliar su uso y a la vez actualizar los resultados obtenidos en el contexto en que se aplica.

La presente investigación se desarrolla en el contexto específico de la evaluación, enfocada al tema de validez, para lo cual el Centro de Investigaciones Educativas de la Universidad del Valle, ha proporcionado los datos obtenidos en diferentes aplicaciones de la prueba de matemáticas de tercer grado, realizadas con diversas muestras de sujetos, que incluyen datos de tercero primaria y de sexto primaria, de establecimientos educativos del país.

II. DEFINICIÓN DEL PROBLEMA

La validez de los resultados que se obtienen por medio de instrumentos de medición, constituye un requisito fundamental dentro de la teoría psicométrica, principalmente debido a que al momento de dar a conocer las consecuencias de las mediciones en cuestión, es importante que ello sea lo más ajustado a la realidad del individuo, en cuanto a las características o cualidades del atributo que se está midiendo. La validez es, sin duda, la consideración más importante en la elaboración y evaluación de los instrumentos de medición, tal como se indica en la edición de 1999 de los Estándares para la Evaluación en Educación y Psicología (Padilla, Gómez, Hidalgo y Muñiz, 2006).

Asimismo, la validez de un *test* o una prueba estandarizada garantiza que la información que se está proporcionando al sujeto de evaluación o al público a quien se reporta su resultado, es la evidencia más cercana de la variable que posee o lo caracteriza. De tal manera que, si la información que se divulga no es “válida”, se estaría representando de manera incompleta o equivocada dicha variable del individuo o de los individuos, y generalmente a lo que conduciría es a tomar decisiones inadecuadas sobre un grupo de personas o sobre un sujeto en especial.

Por lo tanto surge la pregunta, ¿es válida la prueba que utiliza el Centro de Investigaciones Educativas para evaluar el área de matemáticas de tercer grado? Y a partir de ello, se plantea el siguiente objetivo:

Evaluar la validez predictiva, de constructo y del nivel cognitivo de los ítems de una prueba de matemáticas de tercer grado.

III. MARCO TEÓRICO

A. Antecedentes e investigaciones realizadas

Los primeros estándares profesionales en la evaluación surgen en los años cincuenta cuando la American Psychological Association [APA] publica en 1954 un documento sobre las recomendaciones técnicas para los tests psicológicos. El término “Estándares” aparece en la edición de 1966 realizada en forma conjunta por las tres asociaciones (American Educational Research Association [AERA], APA y National Council on Measurement in Education [NCME]) y en 1974, los estándares marcan una nueva etapa en la preocupación por las consecuencias del uso de las pruebas (Padilla *et al.*, 2006) y recomiendan investigar posibles diferencias en la validez relacionada con el criterio para sub-muestras definidas por variables demográficas. La definición de validez de los Estándares de 1985, planteaba que esta “se refiere a la adecuación, significado y utilidad de las inferencias específicas hechas desde las puntuaciones en el test”. En la edición de los Estándares de 1999, se reconoce las ventajas de hacer equivalente la justicia en las pruebas a una validez comparable para todas las personas y se resaltan las aportaciones de Samuel Messick en el debate sobre la incorporación al análisis de la validez de las consecuencias sociales del uso de las pruebas como una fuente más de evidencia dentro del proceso de validación.

Bajo la perspectiva histórica del concepto de validez, en sus inicios se hablaba de validez aparente, validez intrínseca, por definición, validez lógica y validez factorial, entre otros términos (Gómez e Hidalgo, s.f.). Más adelante se delimitaron y denominaron los tipos de validez que una prueba debe incluir, refiriéndose a la validez de contenido, predictiva, concurrente y de constructo. Y en general la validez se ha definido como el grado en el que la prueba mide lo que pretende medir, o sirve para el propósito por el que ha sido construido.

Por aparte, Cronbach y Meehl desde 1955 hablan de estudiar diferencias entre grupos, correlaciones entre subpruebas, estudios sobre la estructura interna de la prueba, estudios factoriales, correlaciones con criterios externos, estudios longitudinales, estudios experimentales y estudios sobre el análisis del contenido. Luego Campbell y Fiske en 1959

introducen los conceptos de validez discriminante y convergente, y la matriz multimétodo-multirrasgo para evaluarlas. Para complementarlo con lo que Messick en 1975 afirma sobre que toda la evidencia de validez debe ser de constructo y contempla un aspecto más referido a los valores sociales y las consecuencias éticas (Gómez e Hidalgo, s.f.).

Asimismo, en la última mitad del siglo XX, se ha investigado una extensa variedad de técnicas de análisis para establecer la validez de una prueba, lo cual implica también obtener evidencia de que el instrumento con el que se trabaja esté libre de sesgo, es decir, los ítems de la prueba funcionan del mismo modo para distintos grupos en función de variables sociodemográficas, cognitivas o de cualquier otro tipo que pueda constituir una fuente sistemática de variación ajena al constructo medido por la prueba (funcionamiento diferencial del ítem).

En cuanto a la evolución de la definición del término validez, se puede mencionar que esta también es específica a otros campos más allá de la educación, por ejemplo, la validez legal, moral, organizacional y técnica entre otras. Asimismo la validez en investigaciones cuantitativas, abarca términos como validez interna, externa, validez de la población y validez ecológica; mientras que en investigaciones cualitativas se hace mención de validez descriptiva, interpretativa, teórica, evaluativa, comunicativa, pragmática, transaccional y aparente, entre otras (Newton, 2013).

Respecto a diversas investigaciones realizadas en torno al tema de validez, en el 2005, Castrillón y Borrero realizaron un estudio con el propósito de estandarizar el Inventario de Ansiedad Estado – Rasgo en niños y adolescentes entre 8 y 15 años de edad, para lo cual evaluaron a 670 niños de ambos géneros, en 35 instituciones educativas públicas y privadas de la ciudad de Medellín, Colombia. El análisis se desarrolló evaluando las propiedades estructurales de la prueba, a través de la validez de constructo hallada a partir del análisis factorial con rotación varimax por el método de componentes principales, y la confiabilidad por alpha de Cronbach con coeficiente de correlación intraclase. El análisis de dichas condiciones arrojó como resultado seis factores en el análisis factorial que explicaron el 51.74% de la varianza. Se estableció un alpha de Cronbach de 0.70. Los resultados de este

trabajo son mucho más cercanos culturalmente al resto de la población colombiana, que a la población española en donde se hizo la adaptación en castellano.

De la misma forma, Barcelata, Gómez y Durán (2009) realizaron una investigación sobre la construcción, confiabilidad, validez de contenido y discriminante del Inventario Autodescriptivo del Adolescente –IADA-, cuyo objetivo es la identificación de factores en las dimensiones familiar, personal, social, escolar, salud y sexual. La construcción del instrumento se realizó en dos estudios, uno de ellos para examinar la validez de contenido a través de los coeficientes de concordancia *kappa* con la participación de 31 jueces expertos, y otro, para evaluar la confiabilidad y validez discriminante. Participaron tres muestras de adolescentes de 13 a 18 años de edad, una muestra piloto de 113 estudiantes, una muestra normal de 1660 estudiantes de secundaria y preparatoria de escuelas públicas y privadas, y una muestra clínica de 76 adolescentes que asistían a instituciones de salud pública y privada de la Ciudad de México. Los análisis mostraron coeficientes de concordancia *kappa* altos (0.935), así como también alfas de Cronbach altos (alpha global 0.90). Se encontraron diferencias significativas con *t de Student*, entre la muestra normal y la clínica, lo cual apoya la capacidad del IADA para discriminar e identificar poblaciones en riesgo.

Por otro lado, González-Forteza *et al.* (2011) efectuaron un estudio sobre la confiabilidad y validez de la escala de depresión CES-D, en un censo de estudiantes de nivel medio superior y superior, en la ciudad de México. Se obtuvo información de 57,403 alumnos, de los cuales 47.5% fueron hombres y 52.5% mujeres, con una edad promedio de 17 años (± 2.7). La escala CES-D se incorporó en los formatos del Examen Médico Automatizado (EMA) que se utilizaron para realizar el levantamiento censal entre los alumnos de nuevo ingreso a bachillerato o licenciatura. La consistencia interna de la escala global mostró un índice satisfactorio ($\alpha=0.83$). El análisis factorial con extracción de componentes principales y rotación oblicua arrojó una estructura de cuatro factores con valores Eigen mayores a uno, que explican 54.1% de la varianza. La validez concurrente se analizó mediante su asociación con la Escala de Expectativas de Vivir-Morir de Okasha, que también se aplica en el EMA, obteniendo una correlación positiva y significativa entre

quienes obtuvieron sintomatología depresiva elevada y quienes manifestaron deseos de morir.

Por aparte, Escartín, Rodríguez-Muñoz, Sora y Rodríguez-Carballeira (2012) llevaron a cabo la adaptación y validación de una escala de conductas negativas realizadas (NAQ-P), basada en la Escala de Conductas Negativas-Revisada para víctimas, de Moreno et al., en el 2007. Con una muestra de 521 empleados de un total de 20 organizaciones de diferentes sectores de actividad ubicadas en España, se recabaron datos acerca de las conductas negativas realizadas y recibidas, así como de otras variables como el liderazgo transformacional, el género o el estatus ocupacional. Los resultados mostraron que un modelo de dos factores (acoso personal -3 ítems- y acoso relacionado con el trabajo -4 ítems-) se ajustaba mejor a los datos que los otros modelos considerados. Además, los resultados mostraron que el instrumento tenía una buena consistencia interna y una adecuada validez de constructo y discriminante.

Al igual, en el 2012, González, López y Montenegro realizaron un estudio sobre el análisis de confiabilidad y validez del instrumento Course Experience Questionnaire (CEQ), empleado para conocer la percepción de los estudiantes sobre la calidad del aprendizaje en la educación superior, que incluye las escalas “buena docencia”, “metas y objetivos claros”, “evaluación apropiada”, “carga de trabajo apropiada” e “E-learning”. El cuestionario fue traducido y aplicado a 325 estudiantes de ingeniería de una universidad pública de la región metropolitana de Chile. Se estimó la confiabilidad a través del Alfa de Cronbach para cada una de las escalas del instrumento, con valores desde 0.52 a 0.85 y para estimar la validez de constructo, se realizó un análisis de componentes principales con rotación Varimax, el cual mostró claramente que los factores corresponden a las escalas del cuestionario CEQ. Asimismo, para analizar la validez concurrente se procedió a correlacionar los resultados del CEQ con los resultados del cuestionario Study Process Questionnaire (SPQ) que se emplea para indagar cómo los estudiantes universitarios abordan el aprendizaje. En cuanto a la validez concurrente se presentaron correlaciones negativas y significativas con la escala “aprendizaje superficial” ($r = -0.318$, $p < 0.01$).

En el 2013, para Cervantes, Álvarez, Gómez y Mancilla el propósito de su investigación fue obtener una versión en español de la Body Modification Scale (BMS), así como examinar sus propiedades psicométricas en varones adolescentes. El primer estudio comprendió la traducción, adaptación y aplicación piloto de la BMS; la consistencia interna y estructura factorial fue evaluada con 270 adolescentes, con una media de edad de 12.84 años, de los cuales 171 participaron en el re-test. El segundo estudio se dirigió a probar, en una muestra independiente de 200 adolescentes, con media de edad de 13.46 años, la adecuación de la estructura factorial derivada respecto a la original. La versión en español de la BMS mostró tener adecuada consistencia interna ($\alpha = 0.88$) y confiabilidad test-retest ($r = 0.80$). El análisis factorial exploratorio derivó inicialmente cuatro factores, observándose traslapamiento conceptual entre dos de ellos; por tanto se optó por replicar el análisis forzando la extracción a tres factores. Estos explicaron 44.1% de la varianza, agrupando 20 de los 24 reactivos originales. En general, la versión en español de la BMS mostró poseer adecuadas propiedades psicométricas en varones adolescentes; sin embargo esto deberá corroborarse no sólo en otras edades, sino también en mujeres.

También, en el 2013, Ramírez, Díaz, Ramos y Manga llevaron a cabo un diseño longitudinal de panel para determinar la validez de la Batería Luria Inicial, con el propósito de identificar alteraciones neuropsicológicas en la edad escolar en una población cubana con atraso en el proceso de adquisición de las funciones psicológicas superiores. La investigación incluyó 126 niños: 89 con trastorno por déficit de atención con hiperactividad, 25 con signos de inmadurez neuropsíquica y 12 con factores de riesgo neurológico por hipoxia perinatal. Se calculó el coeficiente de Cronbach y los resultados indicaron que el instrumento tiene una consistencia interna menor que las normas españolas (0.92), pero mantiene un valor estadístico confiable (0.82). El porcentaje de la varianza explica al 58.64% comparado con las normas españolas, 59.73%. Y en comparación con la versión mexicana queda 62.5% (cubana) y 68.51% (mexicana). Además presenta un fuerte valor predictor de alteraciones ejecutivas después de tres años de aplicado el instrumento (valores r mayores a 0.40).

Asimismo Cruz, Díaz-Loving, Oropeza, Díaz, Padilla y González (2013) realizaron un estudio para mostrar evidencias de validez y confiabilidad de una traducción al castellano del Instrumento de Temor a la Evaluación Social Negativa, en una muestra de estudiantes universitarios de la ciudad de México. Se realizaron análisis de discriminación de reactivos, factoriales exploratorios y factoriales confirmatorios mediante modelamiento de ecuaciones estructurales. Los resultados mostraron una estructura unifactorial que agrupa ocho de los doce reactivos del instrumento original, explica el 53.8% de la varianza total y muestra consistencia interna adecuada ($\alpha = 0.88$).

En el 2014, Márquez-Sandoval *et al.* diseñaron y validaron un cuestionario para valorar el comportamiento alimentario en estudiantes universitarios mexicanos. Se determinó su validez de contenido mediante la evaluación por expertos. Se aplicó en dos ocasiones con un intervalo de cuatro semanas a 333 estudiantes del sexto semestre de Medicina y Nutrición de la Universidad de Guadalajara. Su validez de constructo y consistencia interna se calculó a través del análisis Rasch, tanto para la dificultad de los ítems como para la habilidad de los sujetos. El cuestionario quedó integrado por 31 preguntas con opción de respuesta múltiple. Se obtuvo un alfa de Cronbach de 0.50 para la habilidad de los sujetos y de 0.98 para la consistencia interna de los ítems. El 87.1% de los sujetos y el 89.8% de los ítems tuvieron valores de Infit (calibraciones cercanas al nivel de habilidad del respondiente) y Outfit (calibraciones sensibles a los valores extremos alejados de la habilidad del respondiente) dentro de los valores aceptables (entre 0.8 y 1.2 logits).

Asimismo en 2014, Romo *et al.* realizaron un estudio sobre validación del inventario de ajuste socioemocional de Weinberger en una población de hispanoparlantes mexicanos. El cuestionario traducido del inglés y revisado se aplicó a 452 participantes, respetando la escala Likert del cuestionario original de cinco puntos. Los participantes fueron personas adultas (mayores de 18 años) sin distinción de sexo, que tenían un nivel de escolaridad mínimo de secundaria. Después del análisis de la capacidad discriminadora de los reactivos por medio del análisis factorial de componentes principales, el cuestionario quedó conformado por 44 reactivos, agrupados en cuatro factores: 1. Dominio de sí mismo, con un total de 19 preguntas que explicaron 18.68% de la varianza, 2. Experiencia subjetiva de

distrés, compuesto por 12 preguntas que explicaron 10.84% de la varianza, 3. Defensividad, el cual constó de 7 preguntas que explicaron el 7.10% de la varianza, y 4. Consideración hacia los demás, que agrupó 6 preguntas, que explicaron 6.54% de la varianza. Finalmente se evaluó la confiabilidad del cuestionario utilizando el alpha de Cronbach: Dominio de sí mismo 0.89, Experiencia subjetiva de distrés 0.84, Defensividad 0.69 y Consideración hacia los demás 0.74.

Por su parte Toro, Arias y Quant (2014) diseñaron y validaron el Cuestionario de Autoesquemas Desadaptativos en la Fobia Social para evaluar los contenidos cognitivos disfuncionales de estos trastornos. Se obtuvo la validez de contenido por medio de tres jurados expertos. Con una muestra de 1,413 participantes con edad promedio de 26.41 años ($DE=8.51$), la consistencia interna fue $\alpha=.86$, con una adecuada validez de criterio con la Escala de Temor a la Evaluación Negativa ($r=.656$, $p<.01$). A su vez, según el análisis de la teoría de respuesta al ítem se ajustó a los parámetros del modelo, el cual asume que cada ítem tiene su propia gradación. Al final se recomienda este instrumento para la evaluación de la fobia social durante el tratamiento cognitivo conductual.

De igual forma, en 2014, Omar, Salessi y Urteaga diseñaron y validaron una escala para medir capital psicológico en adultos, es decir aquellas facetas de la personalidad que contribuyen a la productividad de los empleados. Los ítems iniciales se elaboraron sobre la base de la revisión bibliográfica, entrevistas con empleados y consultas con especialistas. Inicialmente se efectuó una prueba piloto que permitió realizar ajustes conceptuales y semánticos de los ítems. La escala fue administrada a 382 empleados de organizaciones públicas y privadas, radicadas en la zona centro de Argentina, conjuntamente con las medidas de desempeño laboral, comportamientos organizacionales contraproducentes y satisfacción general. Asimismo se estudió la validez factorial exploratoria y confirmatoria, además de la confiabilidad como consistencia interna a través del coeficiente alpha de Cronbach. Los resultados mostraron la existencia de cuatro factores con satisfactorios niveles de validez y confiabilidad, los que fueron rotulados como: Esperanza ($\alpha=0.87$), Optimismo ($\alpha=0.91$), Resiliencia ($\alpha=0.84$) y Autoeficacia ($\alpha=0.79$) y que explicaron el 57% de la varianza total.

A nivel nacional, Fortín (1998) realizó un estudio sobre la validez concurrente de los instrumentos elaborados por las maestras de una institución privada mixta de la ciudad de Guatemala, para evaluar el aprestamiento a la lectura. La población estudiada estuvo constituida por 68 niños que iniciaban la segunda parte de kindergarten y 65 niños que iniciaban la preparatoria. Dichas pruebas se aplicaban de manera rutinaria en los meses de junio y octubre para evaluar el funcionamiento de los alumnos. Se realizó una comparación de los instrumentos desarrollados por las maestras de la institución, con instrumentos criterio, como el caso de las pruebas Metropolitan Readiness Tests (MRT) español en sus niveles 1 y 2. La correlación entre las pruebas diseñadas por las maestras y el MRT fueron significativas, por lo que pudo concluirse que los instrumentos desarrollados por las maestras son instrumentos válidos para la evaluación del aprestamiento a la lectura en los alumnos e incluso poseen características que se adecúan mejor para funcionar como un instrumento criterio (objetivos establecidos por las maestras).

Por otra parte Armas, en el año 2002, realizó una investigación sobre la validación de los instrumentos de evaluación docente en la facultad de Ciencias Químicas y Farmacia de la Universidad de San Carlos de Guatemala. La población estudiada fueron profesores y alumnos de dicha Facultad. Los instrumentos utilizados constan de tres áreas: aspecto didáctico y pedagógico (27 preguntas), aspecto psicosocial (25 preguntas) y aspecto profesional (9 preguntas). En total fueron 61 preguntas que se responden en una escala ascendente de 0 a 4. Se obtuvo una alta confiabilidad y se establecieron correlaciones, de acuerdo al tipo de evaluación, para determinar la validez concurrente. Se evidenciaron correlaciones bajas, lo cual refleja que no hay acuerdo entre las diferentes instancias de evaluación aplicando el mismo instrumento, lo que significa que los instrumentos carecen de esta validez.

Asimismo Soto (2007) llevó a cabo una investigación con el objetivo de predecir el éxito o fracaso de estudiantes repitentes de primer año de la Facultad de Ciencias Médicas de la Universidad de San Carlos (USAC). Para tal fin estudió a 970 estudiantes que en el 2003 se inscribieron en primer año en la categoría de repitentes y se incluyeron en el trabajo variables del estudiante registradas en el departamento de Registro y estadística de

la USAC, tales como edad, sexo, título de nivel medio, tipo de establecimiento, lugar de procedencia, instrucción del padre y de la madre, así como su desempeño académico en la Facultad incluyendo el promedio, número de años inscritos en primer año, cursos asignados, índice de aprobación, índice de repitencia y rendimiento académico (este último medido como aprobado o reprobado). Se utilizó regresión logística para la creación del modelo, método estadístico utilizado para predicción de pertenencia a categorías, en el cual intervienen múltiples variables. El modelo quedó construido con las variables que resultaron significativas, siendo estas el índice de repitencia, el promedio y los cursos asignados. La investigación confirmó que estas tres variables predictivas, permiten estimar la probabilidad de éxito del estudiante repitente en el momento de asignarse nuevamente el año perdido, obteniéndose así una visión del acompañamiento que el estudiante requiera y permitir un mayor grado de eficiencia institucional al solo aceptar estudiantes repitentes con determinada probabilidad de éxito.

Por aparte, Gálvez-Sobral (2008) realizó una investigación con el objetivo de desarrollar y validar la escala Bullying para estudiantes de sexto primaria del municipio de Guatemala. La muestra para el estudio estuvo compuesta por 1232 estudiantes de sexto primaria, pertenecientes a establecimientos públicos y privados, de diferentes situaciones socioeconómicas y seleccionados aleatoriamente. Los estudiantes respondieron de forma anónima a la escala, así como los docentes identificaron de forma anónima a los niños que consideraban víctimas, mediante la escala Bullying para docentes. Los resultados indican que la escala muestra buena consistencia interna (alfa de Cronbach 0.878), por lo que se concluye que la construcción de la misma fue correcta y los ítems discriminan bien. En cuanto a la validez interna del instrumento se encontró evidencia estadísticamente significativa entre el grupo de víctimas seleccionado por el docente y el grupo de víctimas identificado por la escala Bullying, mostrando utilidad en su etapa inicial y recomendando su futura investigación.

B. Importancia de la medición en psicología y en educación

La meta de la psicología como ciencia, es la creación de leyes o principios generales para explicar y predecir los fenómenos que estudia, es decir, pretende la descripción,

explicación y predicción del comportamiento de las personas y la medición es el componente esencial en este proceso, ya que constituye una actividad prioritaria en la ciencia psicológica y en especial en la evaluación psicológica (Aragón, 2011)

De acuerdo con Nunnally (citado en Aragón, 2011), el problema más importante para la psicología es la obtención de métodos adecuados de medición, ya que una teoría solo puede ser probada si las variables hipotetizadas pueden ser medidas de manera adecuada, es decir, si se pueden establecer relaciones funcionales entre estas variables.

Una vez que se ha planteado una hipótesis, hay que ponerla a prueba mediante los criterios científicos usuales de la evidencia y la probabilidad y, para poder hacer esto, se necesita un instrumento de medición.

Para Nunally (citado en Aragón, 2011) la medición consiste en un conjunto de normas para asignar números a los objetos de modo tal que estos números representen cantidades de atributos. El propósito es cuantificar y clasificar los atributos de las personas para definir si las propiedades medidas caen en las mismas o diferentes categorías.

Asimismo Nunally (1987) menciona que en la ciencia psicológica y en educación, los objetos generalmente son los seres humanos, pero las personas solo actúan como parte del proceso de medición, ya que son sus respuestas a los instrumentos de medición lo que se desea conocer. Tales respuestas pueden relacionarse con estándares externos, mediante investigaciones empíricas, con el objeto de mejorar la explicación científica de dichas relaciones y explicar fenómenos importantes.

El objetivo por tanto de la medición en psicología es determinar cuánto está presente de un atributo en determinada persona y esto requiere una expresión numérica de la cantidad. En el nivel más bajo de cuantificación, la medición se ocupa de la presencia o ausencia del atributo; en el nivel intermedio, determina si se posee más o menos cantidad del atributo medido, y, en el nivel más alto, qué tanto se posee del atributo y si se pueden cuantificar las diferencias presentadas en él (Aragón, 2011).

1. Medición de los constructos psicológicos. Para Bunge y Ardila (citados en Tornimbeni, Pérez y Olaz, 2008) el problema en psicología, es que muchos constructos, es decir, conceptos teóricos que no son directamente observables, y sus indicadores operacionales no han sido definidos ni explicados claramente por teorías científicas y, por consiguiente, la medición no solo es indirecta, sino meramente empírica y ambigua. De tal cuenta que las pruebas psicológicas se construyen para medir constructos que no pueden observarse directamente.

Al respecto Martínez (citado en Aragón, 2011) establece que la medida comienza con un procedimiento para identificar los elementos del mundo real con los elementos o constructos del sistema lógico, por medio de una definición semántica precisa; una vez hecho esto, se procede a la medición propiamente dicha, identificando el objeto o sujeto que se medirá, identificando la propiedad del objeto o de la conducta que se quiere medir en el sujeto (atributo), seleccionando el instrumento y aplicando una regla de asignación numérica por la que se otorga una cifra a la propiedad de la unidad experimental (sujeto) que se mide.

Al hablar de constructos psicológicos es necesario considerar medidas de diferencias individuales en experimentos controlados, tal es el caso de factores verbales, memoria y aprendizaje. A partir de ello podrían formularse hipótesis acerca de las relaciones entre diversos factores de las habilidades, como sería determinar si la práctica en un tipo de aprendizaje puede facilitar el aprendizaje en una nueva situación, enfocándose en la cantidad y el tipo de aprendizaje que se produciría.

Debido a que hay muchas posibilidades en la investigación de encontrar relaciones entre factores de habilidad y procesos de aprendizaje, es necesario enfocarse en esos factores de habilidad y los instrumentos que los miden, a través de un acercamiento a la teoría psicométrica.

2. Medición psicométrica. La psicometría se ocupa de los procedimientos de medición del comportamiento humano, incluyendo a las denominadas pruebas psicológicas (Tornimbeni *et al.*, 2008).

Y en cuanto al tema de medición psicométrica es importante mencionar los orígenes de las pruebas psicológicas. Como se menciona en Aiken (2003) en el año 2200 a.C. el emperador chino instituyó un sistema de exámenes para evaluar a sus oficiales de gobierno cada tres años y determinar si eran aptos para desempeñar sus labores. Desde la dinastía Chang (1115 a.C.) ya tenían un programa de pruebas para el ingreso de funcionarios públicos que evaluaba destrezas importantes para la época, tales como arquería, equitación, música, escritura y matemática (Cohen y Swerdlik, citados en Tornimbeni *et al.*, 2008).

Durante el siglo XIX, los gobiernos británico, francés y alemán diseñaron sus exámenes para el servicio civil, tomando el antiguo sistema chino como patrón. A finales del siglo XIX se inició la evaluación del estudio científico de las diferencias individuales en cuanto a aptitudes y personalidad. Francis Galton fue pionero de la psicología experimental en la Gran Bretaña del siglo XIX. Se interesó en las bases hereditarias de la inteligencia y en la medición de las habilidades humanas y diseñó varias técnicas para investigar las diferencias individuales en cuanto a aptitudes y temperamento. Demostró que podían diseñarse pruebas objetivas y obtenerse puntuaciones significativas a través de procedimientos estandarizados. James Cattell estudió psicología experimental junto con Wundt como con Galton. Se interesó en las diferencias individuales e intentó relacionar las calificaciones de las mediciones de tiempo de reacción y discriminación sensorial con las calificaciones escolares (Aiken, 2003). Y en 1890 acuña el término “test mental”.

Asimismo Binet y Simon (citados en Aiken, 2003) desarrollaron una herramienta práctica dirigida a evaluar casos de niños especiales. Construyeron la primera prueba mental que contribuyó en forma significativa a la predicción del aprovechamiento académico, que se publicó en 1905 y así surgió la primera escala formal para evaluar la inteligencia en niños.

Otros pioneros en pruebas y evaluaciones psicológicas fueron Spearman en teoría de las pruebas, Thorndike en pruebas de aprovechamiento, Terman en pruebas de inteligencia, Woodworth y Rorschach en pruebas de personalidad, Strong en mediciones de interés y Otis con pruebas de inteligencia administradas colectivamente (Aiken, 2003).

La Segunda Guerra Mundial, también estimuló la construcción de pruebas de aptitudes específicas, que fueron muy útiles para seleccionar pilotos, bombarderos, operadores de radio y otras funciones militares especializadas. El Test de Aptitudes Diferenciales, entre otros, surge a partir de esos descubrimientos (Tornimbeni *et al.*, 2008).

En la década de 1950 aparecieron textos que con el tiempo serían clásicos y dejarían establecidos los fundamentos teóricos de la psicometría. Surge en ese momento histórico una corriente de revisión y análisis de la fundamentación científica de las pruebas y los trabajos realizados en este período versan en su gran mayoría sobre teoría de la medición, los principios y fundamentos de la medición en psicología, los problemas de validez y confiabilidad y en sí, la construcción de una teoría psicométrica (Tornimbeni *et al.*, 2008).

En los últimos años del siglo XX se produjo un acercamiento entre la psicometría y la psicología cognitiva y el uso de las computadoras en psicometría tuvo un notable incremento desde la década de 1980, aplicándose en casi todas las instancias de la evaluación psicológica (Tornimbeni *et al.*, 2008).

3. Instrumentos de medición. De acuerdo a Alvarado y Santisteban (2006) la teoría de la medición intenta describir, categorizar y evaluar la calidad de las medidas, mejorar su utilidad, exactitud, su significado, así como proponer métodos para que se desarrollen nuevos y mejores instrumentos de medida.

Los instrumentos de medición deben intentar medir con precisión y considerar que siempre contienen un término de error, debido a que miden una muestra representativa de un constructo. Asimismo los constructos no pueden definirse aisladamente en términos de

definiciones operacionales únicamente, sino que deben establecer relaciones con otros constructos y con otros fenómenos observables (Aragón, 2011).

De esta forma, el valor que se obtiene una vez aplicada una prueba, es el valor empírico mezclado con el error cuya cuantía se desea estimar. Si el instrumento está bien construido, se esperaría que la puntuación obtenida por el sujeto reflejara lo mejor posible su puntuación verdadera y el error de medición fuera cercano a cero; de acuerdo con la teoría clásica de las pruebas, se esperaría que en la medición de atributos, la puntuación obtenida por un sujeto en una prueba se acerque lo más posible a su ejecución real con el mínimo posible de errores, es decir que el instrumento sea confiable para medir el atributo que interesa medir (Aragón, 2011).

Y para poder considerar a un instrumento de medición como adecuado y científico, debe contar con dos requisitos indispensables: la confiabilidad y la validez. En el caso del presente estudio, se enfatiza el tema de validez relacionado a la prueba de matemáticas de tercer grado.

C. Validez

1. Concepto de validez. La definición de validez citada en la última edición de los Estándares (edición de 1999 de la AERA, APA y NCME) se refiere a “un juicio integrado del grado en el que la evidencia empírica y la racionalidad teórica apoyan la suficiencia y conveniencia de las inferencias y acciones basadas en las puntuaciones en las pruebas u otros modos de evaluación” (Padilla *et al.*, 2006).

De acuerdo a Morales (2011) la validez no es una propiedad del instrumento de medición, sino de las inferencias e interpretaciones que se hacen con los datos obtenidos.

Según menciona Martínez (citado en Aragón, 2011) el estudio de validez de un instrumento se refiere básicamente a validar los datos proporcionados por éste: el grado de adecuación, significación y utilidad de las inferencias específicas que pueden derivarse a partir de las puntuaciones de las pruebas. Es decir, lo que se valida no es el instrumento,

sino la interpretación de los datos obtenidos por medio de un procedimiento específico y las inferencias que se pueden deducir de la ejecución de un sujeto en una prueba. En este sentido, la validación es un proceso permanente, por lo cual no se valida un instrumento de medición, sino el uso que se le da.

Relacionado a lo anterior, el término “validez” se refiere al grado en el cual las conclusiones o interpretaciones derivadas de los resultados de cualquier medición están “bien fundamentadas o justificadas, siendo a la vez relevantes y significativas” (Cook y Beckman, 2006).

De igual forma en el artículo de Gómez e Hidalgo (s.f.) la validez es vista como un argumento cuidadosamente estructurado, que reúne evidencia desde una variedad de recursos para apoyar o refutar las interpretaciones propuestas de las puntuaciones de un instrumento. Los resultados de cualquier medición psicométrica tienen significado o validez solamente en el contexto del constructo que se propone medir. Por lo tanto, la validez no es una propiedad del instrumento, sino una interpretación de los datos obtenidos por un determinado procedimiento, por lo que la validez de las puntuaciones de la prueba debe ser establecida en cada uso que se haga del mismo. Por ello se puede decir que la validez no puede ser nunca comprobada, ya que se convierte en un ciclo de continua evaluación y revisión (Cook y Beckman, 2006).

Al respecto Messick (citado en Gómez e Hidalgo, s.f.) afirmaba que la validez no es de la prueba o de la observación; lo que se valida son las inferencias derivadas de las puntuaciones de la prueba o de otros indicadores, inferencias sobre el significado de las puntuaciones o la interpretación para propósitos aplicados y sobre las implicaciones para la acción, es decir, las consecuencias sociales y éticas.

Asimismo Messick (citado en Cook y Beckman, 2006) señala como fuentes de evidencia a considerar en la validez:

- Contenido: relevancia y representatividad del “constructo” de la prueba.

- Sustantivo: razones teóricas de la consistencia observada de las respuestas (acciones y procesos de pensamiento). Aquí se puede preguntar a un grupo de estudiantes que “piensen en voz alta” mientras responden un conjunto de preguntas.
- Estructural: configuración interna de la prueba (confiabilidad) y dimensionalidad (análisis factorial); funcionamiento diferencial de los ítems.
- Generalización: grado en que las inferencias hechas a partir de la prueba se pueden generalizar a otras poblaciones, situaciones o tareas. Este aspecto tiene especial importancia en la adaptación y/o traducción de escalas y pruebas de una cultura a otra.
- Externo: relaciones de la prueba con otras pruebas y constructos. Análisis de la utilidad de la medida.
- Consecuencial: consecuencias éticas y sociales de la prueba. Evaluación del sesgo de la prueba.

Por aparte Kane en el año 2001 (citado en Padilla, Gómez, Hidalgo y Muñiz, 2006) resalta los cuatro aspectos más notables del consenso actual sobre la teoría de la validez: a) la validez implica una evaluación de la plausibilidad de la interpretación y del uso propuesto de las mediciones; b) los juicios de validez reflejan la suficiencia y conveniencia de la interpretación propuesta y de las evidencias que la apoyan; c) la medición debe incluir la evaluación de las consecuencias del uso de las pruebas; y d) la validez es una evaluación unificada de la interpretación propuesta.

2. Fuentes de validez. La clasificación más importante de las fuentes de validez es la creada por el comité conjunto de la Asociación Americana de Psicología, la Asociación Americana de Investigación Educativa y el Consejo Nacional de Mediciones en Educación. En ella se incluyen tres fuentes de validez: de contenido, relacionada con el criterio y de constructo (Kerlinger y Lee, 2002).

De acuerdo a Cohen y Swerdlik (2006) son tres los enfoques para evaluar la validez asociada respectivamente con la validez de contenido, la validez relacionada con el criterio y la validez de constructo:

- Examinar el contenido de la prueba.

- Relacionar las calificaciones obtenidas en la prueba con otras puntuaciones u otras medidas.
- Realizar un análisis general de:
 - La forma en que las puntuaciones de la prueba se relacionan con otras medidas y calificaciones (otras pruebas).
 - La forma en que las puntuaciones de la prueba pueden ser entendidas dentro de un contexto teórico para comprender el constructo a medir y por el cual la prueba fue diseñada.

Según dichos autores estos tres enfoques sobre la validez de la evaluación no son mutuamente excluyentes; cada uno debe ser considerado como un tipo de evidencia que, junto con otras, contribuye a elaborar un juicio sobre la validez de la prueba.

a. Validez referida al criterio. La validez relacionada con el criterio se estudia al comparar las puntuaciones de una prueba o escala con una o más variables externas, o criterios, que se sabe o se considera que miden el atributo que se estudia. Para obtener el valor de la validez del criterio, se correlacionan los puntajes de la prueba con los puntajes de la variable criterio (Aragón, 2011).

Para Cohen y Swerdlik (2006) la validez relacionada con el criterio es un juicio de cuán adecuadamente puede ser utilizada la puntuación de una prueba para inferir la posición más probable de un individuo con respecto a cierta medida de interés, siendo el criterio esa medida de interés. Un criterio puede ser definido como el modelo contra el cual se compara y evalúa una prueba o la puntuación de una prueba.

Según menciona también Aragón (2011) la validez referida al criterio se caracteriza por la predicción relacionada con un criterio externo y porque recurre a la comprobación del instrumento de medición, ya sea en el momento presente o en el futuro, y lo compara con un resultado o medida. De acuerdo a esto, la validez relacionada con el criterio está caracterizada por la predicción sobre un criterio externo y por la verificación de un instrumento de medición, ya sea ahora o en el futuro, contra un resultado o medida; aunque

muchas veces la mayor dificultad es el criterio mismo, es decir, es difícil obtener un criterio “correcto” o adecuado para llegar a comprobar la validez.

De acuerdo a Kerlinger y Lee (2002) un tipo de validez relacionada con el criterio es la llamada validez predictiva; el otro tipo es la validez concurrente, que difiere de la predictiva en la dimensión del tiempo. La validez predictiva involucra el uso de desempeños del criterio futuros y está determinada por el grado de correspondencia entre las medidas implicadas; mientras que la validez concurrente, mide el criterio casi al mismo tiempo y se calcularía al correlacionar dos conjuntos de calificaciones. Relacionado a esto, la validez concurrente indica el grado en el que las puntuaciones de una prueba pueden servir para estimar la posición actual de un individuo frente a un criterio (Cohen y Swerdlik, 2006).

La validez predictiva tiene mayor importancia en las pruebas escolares que sirven para determinar los métodos de estudio idóneos. En las escuelas, este tipo de validez se emplea en las medidas de “habilidad” (Nunnally, 1987). La validez de los instrumentos predictores se determina por medio del análisis correlacional y sus posibilidades de uso en el análisis multivariado.

Al considerar las ventajas de las pruebas predictoras, es erróneo pensar en términos de correlaciones perfectas, o incluso en muy altas correlaciones. En la mayoría de los problemas de predicción es razonable suponer solo correlaciones modestas entre un criterio y cualquier prueba predictora individual, o combinación de estas. Las pruebas sólo guardan modestas correlaciones con sus criterios (por ejemplo, correlaciones entre 0.30 y 0.40) (Nunnally, 1987).

Al respecto, Cronbach y Gleser (citados en Cohen y Swerdlik, 2006) argumentan que los coeficientes de validez no necesitan ser demasiado elevados para permitir al usuario tomar decisiones precisas dentro del contexto único en el cual una prueba está siendo usada. Por lo que el coeficiente de validez debería ser lo bastante alto para que permita la identificación y diferenciación de los examinados con respecto a un atributo.

b. Validez de contenido. La validez de contenido es la representatividad o la adecuación de muestreo del contenido –la sustancia, la materia, el tema- de un instrumento de medición. Una prueba con alta validez de contenido sería teóricamente una muestra representativa del universo de contenido de la propiedad que se mide. Si fuera posible elegir aleatoriamente reactivos de este universo en número suficiente, entonces cualquiera de estas muestras de reactivos supuestamente formaría una prueba con una alta validez de contenido (Kerlinger y Lee, 2002). De ser posible el universo de contenido debe estar claramente definido; es decir se deben facilitar las instrucciones específicas a los jueces para realizar juicios, así como las especificaciones sobre lo que están juzgando.

Lo que interesa en este tipo de validez es si las respuestas dadas a los reactivos de estímulo y las condiciones bajo las cuales la conducta es observada, representan a todos aquellos conjuntos de estímulos, respuestas y condiciones, a los cuales se interesa generalizar. La validez de contenido tiene como objetivo demostrar que los reactivos de la prueba son una muestra representativa de un universo y asegurar un muestreo cuidadoso de un dominio de contenido relevante (Aragón, 2011).

Hoste (citado en Aragón, 2011) menciona que no existe un modo aceptable para cuantificar la validez de contenido y, por tanto, no puede ser expresada como un coeficiente de validez, por lo que se determina por medio de procedimientos racionales en vez de empíricos, con un examen cuidadoso de los procedimientos de construcción de la prueba. Por lo general, se utiliza la técnica de “jueceo”, en donde jueces competentes en el atributo que se pretende medir y en el objetivo de la prueba, ponderan el contenido de los reactivos (Kerlinger, citado en Aragón, 2011).

Respecto a las pruebas de rendimiento educativo, es usual considerar una prueba como una medida de contenido válido cuando la proporción del material cubierto por la prueba se aproxima a la proporción del material que se cubrió en el curso (Cohen y Swerdlik, 2006).

c. Validez de constructo. Un constructo es una idea informada, científica, desarrollada como una hipótesis para describir o explicar el comportamiento; asimismo los

constructos son rasgos inobservables, supuestos (subyacentes) a los que un desarrollador de pruebas puede recurrir para describir el comportamiento de la prueba o el desempeño del criterio evaluado (Cohen y Swerdlik, 2006).

El propósito de la validez de constructo es el de validar la teoría subyacente al sistema de evaluación y a la medida misma. Los constructos son los que determinan qué conductas han de seleccionarse para su observación y la situación en la que esta será aplicada. Un instrumento de medida estará ligado al sistema de constructos dentro del que fue construido; este tipo de validez indicará el grado en que el instrumento de evaluación es una medida adecuada del constructo y en qué grado o alcance también las hipótesis derivadas del mismo pueden confirmarse mediante la utilización del instrumento en cuestión (Aragón, 2011).

Desde sus primeros estudios, Messick (1975) menciona que la validez de constructo es un requisito esencial de la medición, particularmente en las ciencias sociales y del comportamiento. Asimismo señala que las pruebas en sí mismas no tienen validez de constructo, ni confiabilidad o validez predictiva; sino que estas son propiedades de las respuestas de las mismas, en función de las personas y el contexto de la medición. Asimismo indica que todas las mediciones deben ser referidas al constructo, ya que una medida debe estimar “cuánto de algo” un individuo posee.

De acuerdo a Martínez (citado en Aragón, 2011) la validez de constructo es la extensión en la cual la prueba dice medir un constructo o rasgo teórico y requiere la acumulación gradual de información de diferentes fuentes; cualquier dato que arroje luz sobre la naturaleza del constructo bajo consideración y de las condiciones que afectan su desarrollo y sus manifestaciones, representa una evidencia apropiada para este tipo de validación.

Por lo general, el investigador mantiene una firme creencia en algunas de las observaciones más importantes relacionadas con el constructo, pero más allá de esto solo puede adivinar hasta qué límite se extiende el constructo, ya que los límites del dominio de observaciones relacionadas no son claros.

Asimismo, los investigadores o los expertos en medición, al investigar la validez de constructo de una prueba, buscan explicar diferencias individuales en las puntuaciones de la prueba y su interés está centrado en las propiedades que se miden, más que en las pruebas utilizadas para lograr la medición y por ello debe intentarse validar la teoría que está detrás de la prueba.

De acuerdo a Nunnally (citado en Aragón, 2011), el procedimiento de medición de constructos tiene tres aspectos principales:

- La especificación del dominio de observaciones, es decir extraer el dominio de las observaciones relacionadas con el constructo.
- La determinación del grado en que todas o algunas de estas observaciones se correlacionan entre sí, es decir el grado en que las variables del dominio tienden a medir la misma cosa.
- La determinación de cuándo algunas o todas las medidas de tales variables actúan como si midiesen el constructo. Es decir, si la supuesta medida de un constructo se correlaciona adecuadamente con las medidas de otros constructos.

Por otra parte, según Campbell y Fiske (citados en Messick, 1975) la validez de constructo tiene dos requisitos esenciales. Uno es la evidencia convergente, por la cual se demuestra que una medida en cuestión está sustancialmente relacionada con otras medidas del mismo constructo y otras variables con las cuales debe correlacionar sobre su base teórica. La evidencia de la validez de constructo de una prueba particular puede converger de diversas fuentes, como otras pruebas o medidas diseñadas para evaluar el mismo constructo; de esta manera, si las calificaciones en la prueba pasan por una validación de constructo tenderán a correlacionarse de forma elevada en la dirección predicha, con las calificaciones de pruebas anteriores, más establecidas y ya validadas, diseñadas para medir el mismo constructo, lo cual sería un ejemplo de evidencia convergente (Cohen y Swerdlik, 2006).

Igualmente importante es la evidencia discriminante, por la cual se demuestra que la medida no está relacionada con los indicadores de otros constructos distintos (Campbell y Fiske, citados en Messick, 1975). Un coeficiente de validez que muestra poca relación entre las calificaciones de la prueba y otras variables con las que las puntuaciones en la prueba que se está sometiendo a validez de constructo no debería correlacionarse desde el punto de vista teórico, proporciona evidencia discriminante de la validez de constructo (Cohen y Swerdlik, 2006).

Campbell y Fiske (citados en Messick, 1975) afirman que para obtener evidencia completa de la validez de constructo, las correlaciones deben seguir un patrón establecido y si no se logran cubrir los requisitos de ese patrón, se debilitan los aspectos de la validez. Por tanto, se espera un constructo que esté bien definido en términos de una variedad de observaciones, para las cuales hay una o más variables que representan adecuadamente su dominio, y eventualmente se correlaciona con otros constructos de interés.

Cronbach en 1990 (citado en Kerlinger y Lee, 2002) indica que existen tres partes en la validación de constructo: sugerir qué constructos posiblemente explican el desempeño de la prueba, derivar hipótesis a partir de la teoría que incluye al constructo y comprobar empíricamente las hipótesis. Por lo anterior, la validez de constructo abarca matrices de correlación (método multirrasgo-multimétodo), índices de consistencia interna, análisis factorial y estudios de diferencias entre grupos.

Relacionado a lo anterior, la evidencia convergente y discriminativa de la validez de constructo puede obtenerse por medio del uso del análisis factorial, diseñado para identificar factores o variables específicas que de manera particular son atributos, características o dimensiones en las que es posible diferir (Cohen y Swerdlik, 2006).

D. El sistema cognitivo

1. La taxonomía de Marzano. La taxonomía de objetivos educativos de Marzano y Kendall (2008), surge con un enfoque diferenciado de la taxonomía de Bloom publicada en 1956. De acuerdo a estos mismos autores, Bloom señalaba que un objetivo educativo debe

contener una referencia clara de un tipo específico de conocimiento, así como los comportamientos que son la señal de un entendimiento o habilidad relacionados a dicho conocimiento.

Según Marzano y Kendall (2008), la taxonomía es bidimensional; una dimensión hace referencia a tres dominios de conocimiento, y la otra a niveles de procesamiento cognitivo. Y por conocimiento se entiende aquella evidencia que el estudiante puede dar sobre lo que recuerda, ya sea por evocación o reconocimiento, acerca de un fenómeno o experiencia dentro del proceso educativo.

La taxonomía por lo tanto, postula tres dominios de conocimiento que operan sobre la base de tres sistemas de pensamiento y los elementos que componen estos últimos. Y son los sistemas de pensamiento los que tienen una estructura jerárquica, que constituyen la Nueva Taxonomía.

Los tres dominios de conocimiento son: información, procesos mentales y procesos psicomotores; y las tres categorías de procesamiento son cognitiva, meta-cognitiva y conciencia del ser.

2. Componentes del sistema cognitivo. En la taxonomía se hace mención de dos dimensiones. La dimensión de los dominios del conocimiento: dominio de la información, dominio de los procesos mentales y el dominio de los procesos físicos o psicomotores. Asimismo la dimensión de los niveles de procesamiento: sistema cognitivo, sistema meta-cognitivo y sistema de conciencia del ser.

3. Dimensión de los dominios del conocimiento. Según Marzano (2001), los dominios del conocimiento son: información, procesos mentales y procesos psicomotores.

El dominio de la información, es referido también como aquellos conocimientos declarativos y tiene también una estructura jerárquica: términos de vocabulario, hechos,

secuencias en el tiempo, así como principios y generalizaciones en cuanto a la organización de ideas.

El dominio de los procesos mentales, también referido como conocimiento de los procesos, puede describirse como el “cómo hacer para” e incluye procesos y macro-procesos, así como habilidades sobre tácticas, algoritmos y reglas simples.

El dominio de los procesos psicomotores está compuesto por procesos físicos y está dividido en procesos complejos y procesos simples.

4. Dimensión de los niveles de procesamiento cognitivo. Se abarcan los sistemas de cognición, meta-cognición y conciencia del ser, propuestos por Marzano (2001). Asimismo el sistema de cognición envuelve los procesos de conocimiento/recuerdo, comprensión, análisis y utilización del conocimiento. Y de esta manera se conforma una taxonomía de seis niveles.

Se incluyen procesos como nombrar, ejecutar, sintetizar, representar, relacionar, clasificar, analizar, generalizar, especificar, tomar decisiones, resolver problemas, investigar, especificar metas, evaluar las emociones y la motivación, entre otros (Marzano y Kendall, 2008).

a. Conocimiento/recuerdo. Implica transferir lo que se sabe a un estado actual de atención, tanto en forma de reconocimiento (nombrar), como de evocación o recuerdo (ejecutar). El reconocimiento implica saber si la información es precisa, imprecisa o desconocida, mientras que el recuerdo requiere no solo cierto nivel de reconocimiento, sino además la producción de información almacenada (Marzano y Kendall, 2008).

b. Comprensión. Este proceso es el responsable de identificar atributos críticos y definidos del conocimiento e incluye dos procesos relacionados: integración (síntesis) y simbolización (representación). La *síntesis* envuelve extraer el conocimiento hacia características clave organizadas en una forma generalizada (información y eventos

generales). Y la *representación* es el proceso de crear una analogía simbólica del conocimiento que ha sido producido por el proceso de integración y se da generalmente en forma de imágenes (ej.: organizadores gráficos, símbolos, gráficas) (Marzano y Kendall, 2008).

c. Análisis. Implica la generación de nueva información que el individuo previamente no poseía. Existen cinco procesos de análisis: relación, clasificación, análisis de errores, generalización y especificación. La *relación* se refiere a identificar similitudes y diferencias entre los componentes del conocimiento. La *clasificación* se refiere a organizar el conocimiento en categorías significativas. El *análisis de errores* implica la determinación de falta de lógica y precisión en el conocimiento. La *generalización* es el proceso de construir nuevas generalizaciones desde la información que ya es sabida u observada. Y la *especificación* es el proceso de generar nuevas aplicaciones de una generalización o principio conocidos (Marzano y Kendall, 2008).

d. Utilización del conocimiento. Este proceso es utilizado por el individuo cuando desea completar una tarea específica. En la taxonomía se identifican cuatro categorías de utilización del conocimiento: toma de decisiones, resolución de problemas, experimentación e investigación.

El proceso de *toma de decisiones* es utilizado cuando el individuo debe seleccionar entre dos o más alternativas; los elementos clave de este proceso incluyen la identificación de alternativas, identificar los criterios que se usan al juzgar el valor de cada alternativa y por último, seleccionar entre las alternativas basadas en la aplicación del criterio.

El proceso de *resolución de problemas* es utilizado cuando un individuo intenta lograr una meta, para la cual existen obstáculos; los atributos de este proceso incluyen: identificar los obstáculos hacia la meta, identificar las alternativas para alcanzar la meta, evaluar las alternativas y seleccionar y ejecutar las alternativas.

La *experimentación* es el proceso de generar y comprobar hipótesis para el propósito de entender algunos fenómenos físicos o psicológicos; los atributos críticos de la experimentación incluyen: hacer predicciones basadas en principios conocidos o hipotéticos, diseñar alguna forma para comprobar las predicciones utilizando datos observables y, por último, evaluar la validez de los principios basándose en los resultados de la prueba.

Y finalmente, la *investigación* es el proceso de generar y comprobar hipótesis acerca de eventos pasados, presentes o futuros; los atributos críticos de la investigación abarcan: identificar qué es conocido de acuerdo al fenómeno de investigación, identificar áreas de confusión acerca del fenómeno, proveer una respuesta para la confusión o controversia utilizando opiniones y argumentos expresados por otros y por último, presentar un argumento lógico para la respuesta propuesta (Marzano y Kendall, 2008).

e. Metacognición. Es el responsable de monitorear, evaluar y regular las funciones de todos los otros tipos de pensamiento. Dentro de la taxonomía, el sistema de metacognición tiene cuatro funciones: especificar metas, monitorear el proceso, monitorear la claridad y monitorear la exactitud.

La función de *especificar metas* es responsable de establecer metas claras de aprendizaje para tipos específicos de conocimiento. El componente de *monitorear el proceso*, lo hace sobre la efectividad de un procedimiento que se está llevando a cabo cuando una meta específica se ha identificado. El proceso de *monitorear la claridad* implica determinar hasta dónde el individuo está libre de ambigüedades acerca del conocimiento. Y finalmente, *monitorear la exactitud* envuelve determinar la extensión a la cual un individuo está correcto en términos de entender un conocimiento específico (Marzano y Kendall, 2008).

f. Conciencia del ser. Determina si el individuo se involucra o se aleja de una tarea; y también cuánta energía invierte en la tarea. Existen cuatro tipos de conciencia del ser:

examinar la importancia, examinar la eficacia, examinar la respuesta emocional y examinar la motivación.

El *examinar la importancia* está en función de dos condiciones: es importante porque satisface una necesidad básica o porque ayuda a alcanzar una meta personal. *Examinar la eficacia* involucra hasta qué punto el individuo cree que tiene la habilidad, el poder o los recursos necesarios para alcanzar una competencia relativa a un componente específico de conocimiento. *Examinar la respuesta emocional* involucra analizar hasta dónde el individuo tiene una respuesta emocional hacia un componente específico de conocimiento y la parte que la respuesta juega en la propia motivación. *Examinar la motivación* en general abarca los tres procesos anteriores y es el proceso de identificar el nivel propio de motivación para aprender o incrementar la competencia en un componente dado del conocimiento y luego identificar la interrelación entre la propia creencia de la importancia, la eficacia y las respuestas emocionales que gobiernan el nivel de motivación de sí mismo (Marzano y Kendall, 2008).

En resumen, la nueva taxonomía abarca la dimensión del conocimiento y la dimensión del procesamiento cognitivo. Al utilizar estas dos dimensiones los educadores pueden diseñar objetivos para componentes específicos de conocimiento que involucren procesos cognitivos específicos.

E. Pensar en voz alta (think aloud) como método de validez cognitiva

Thelk, Hoole y Lottridge (s.f.) en su artículo sobre validez cognitiva de ítems y puntuaciones en las pruebas, mencionan que el procedimiento “pensar en voz alta” (think aloud) permite establecer la validez cognitiva de ítems utilizados para medir un constructo. La validez cognitiva es de importancia cuando el constructo de interés es demostrado a través del uso de habilidades o procesos cognitivos.

La validez cognitiva se relaciona directamente con lo que Messick afirmaba sobre la evidencia para la validez sustantiva, la cual se refiere a las razones teóricas de la

consistencia observada de las respuestas (acciones y procesos de pensamiento), al basarse en las habilidades utilizadas mientras se responden los ítems de una prueba.

Quienes desarrollan y utilizan las pruebas necesitan verificar el proceso cognitivo que realizan quienes responden a los tests. Es decir, se trata de una investigación sobre la conexión entre la cognición, la observación y la interpretación. La validez cognitiva es entonces investigada a través del uso de protocolos verbales o evidencia sobre los métodos que utilizan las personas cuando toman una prueba, al momento de resolver las tareas que ello implica (Thelk, Hoole y Lottridge, s.f.).

El método utilizado es que por cada ítem que se presenta, al estudiante se le pide que lea el ítem en voz alta y que explique también en voz alta mientras resuelve el problema. Generalmente estas sesiones son grabadas y luego transcritas para su interpretación y de esta forma llegar a determinar su “alineación” entre la teoría, los expertos y las respuestas de los estudiantes.

El uso de la estructura de validez cognitiva sirve para comparar la demanda de las tareas en las pruebas, la actividad cognitiva actual y la habilidad de las puntuaciones para diferenciar la cualidad cognitiva y es importante, aunque poco utilizado como componente en el desarrollo de las pruebas.

Al respecto Armengol (2007) señala que los protocolos de pensamiento en voz alta son instrumentos metodológicos que implican, como su nombre lo indica, el uso de informantes “pensando en voz alta” mientras llevan a cabo una actividad. Los pensamientos articulados siguiendo esta técnica se graban para poder ser transcritos y son analizados con la ayuda de unas categorías preestablecidas para reflexionar sobre los objetivos del trabajo de investigación. Esta metodología ha sido ampliamente utilizada para analizar actividades de resolución de problemas en numerosos estudios de psicología cognitiva.

Según menciona la misma autora, la psicología cognitiva se basa en la premisa de que los procesos mentales y las habilidades generativo-creativas de los aprendices son

analizables. De acuerdo a Ericsson y Simon (citados en Armengol, 2007) los métodos que utilizan información articulada por informantes pueden ser considerados como uno de los muchos tipos de observaciones que proporcionan datos sobre los procesos cognitivos.

Para estos autores, los informes orales (verbal reports) se clasifican en tres niveles condicionados por aquello que el procedimiento impone en los informantes. Los informes del nivel 1 provienen de la vocalización directa del pensamiento disponible en la memoria de trabajo del sujeto. Los del nivel 2 reflejan descripciones o explicaciones sobre el contenido del pensamiento presente en la atención del sujeto, sin ninguna intención de incitar la búsqueda de información en la memoria a largo plazo. Y finalmente los informes agrupados dentro del nivel 3 resultan de la activación de procesos intermedios para localizar información que no está presente en la memoria de trabajo y, en consecuencia, requieren diferentes grados de introspección.

De acuerdo a Ericsson y Simon (citados en Armengol, 2007), las instrucciones simples que solo implican “pensar en voz alta” producen informes del nivel 1, que no comportan ningún tipo de esfuerzo extra por parte de los sujetos. Las instrucciones que solicitan detalles complementarios (por ejemplo articular todo aquello que pasa por la mente) originan informes del nivel 2.

Los protocolos de pensamiento en voz alta utilizados para analizar actividades de resolución de problemas, se considera que producen informes del nivel 2 y la metodología implica el uso de sujetos “pensando en voz alta” de manera continua mientras llevan a cabo una actividad (Armengol, 2007).

De acuerdo a Casas y Luengo (2004), conocer y representar adecuadamente cómo se producen los procesos de transmisión y adquisición del conocimiento y cómo se construye y modifica la estructura cognitiva del alumno durante el aprendizaje, es decir el estudio de estos procesos mentales, ha sido abordado desde diferentes enfoques y con diversas técnicas, tales como el uso de protocolos de pensamiento en voz alta y entrevistas a profesores y alumnos, entre otros. Aunque también mencionan que la aplicación de estas

técnicas exige un elevado nivel de introspección por parte del alumno, que no siempre es posible en edades tempranas y por último una inferencia, en algunos casos excesiva, por parte del investigador, a la hora de analizar los resultados.

De igual forma, psicólogos y lingüistas han utilizado la técnica de pensamiento en voz alta para incrementar el conocimiento relacionado con este proceso cognitivo tan complejo y, de esta manera, poder desarrollar modelos de los procesos de composición y/o realizar investigaciones empíricas relacionadas con distintos aspectos pedagógicos.

Tal es el caso del estudio realizado por McCorry, Scullion, McMurray, Houghton y Dempster (2013) sobre validez de contenido del cuestionario revisado de percepción de enfermedades (IPQ-R), en pacientes con diabetes tipo 2. El objetivo fue determinar si las cogniciones son congruentes con el significado de los ítems y subescalas, tal como las interpretan los investigadores y clínicos, que utilizan el cuestionario para identificar la naturaleza y extensión de los problemas que los individuos experimentan cuando lo completan. Participaron 36 pacientes de clínicas diabéticas en un hospital y centro de cuidado primario y se les pidió que completaran el cuestionario utilizando el método “pensar en voz alta”. Las transcripciones fueron analizadas para identificar instancias en donde los participantes expresaban problema con la complementación de los ítems, o cuando había inconsistencia entre respuestas verbales y escritas. En general los participantes encontraron el cuestionario sin mayor problema, sin embargo tuvieron problema con los conceptos “cura” y “síntomas” en el contexto de la diabetes tipo 2 y con las frases negativas utilizadas en algunos ítems. Estos hallazgos tienen importantes implicaciones para la interpretación de las puntuaciones del cuestionario, particularmente cuando es utilizado como base para intervenciones individuales en personas con diabetes tipo 2.

Otro estudio relacionado es el de Renberg, Lindblad y Tully (2008) en el que exploraron la validez de contenido y aparente, incluyendo aspectos cognitivos al responder el instrumento de calidad de vida relacionado a la terapia farmacéutica (PTRQoL), traducido del inglés al sueco. Se solicitó a una muestra de 16 clientes de farmacias de una

comunidad sueca que llenaran el instrumento mientras reportaban cómo razonaban en voz alta. Las respuestas fueron grabadas, transcritas literalmente y analizadas. Los respondientes se negaron a leer las instrucciones y variaron en la interpretación de los ítems. La combinación de los protocolos de “pensamiento en voz alta” y el análisis cualitativo informaron sobre la validez del instrumento y fue valioso en el desarrollo del cuestionario. El estudio también identificó problemas específicos que pueden ser relevantes para otros instrumentos para probar actitudes y comportamientos de los pacientes relacionados a la medicina.

IV. MARCO METODOLÓGICO

A. Objetivo general

Evaluar la validez predictiva, de constructo y del nivel cognitivo de ítems de la prueba de matemáticas de tercer grado que aplica el Centro de Investigaciones Educativas, para contar con estudios de validez que aumenten el rigor científico de las pruebas que utiliza.

B. Objetivos específicos

- Evaluar la capacidad predictiva de la prueba de matemáticas de tercer grado a través de la correlación entre los resultados obtenidos en ella y en la prueba de matemáticas de sexto grado, como evidencia de validez predictiva.
- Evaluar la estructura factorial de la prueba de matemáticas de tercer grado y compararla con las especificaciones de la prueba, como evidencia de validez de constructo.
- Evaluar el nivel cognitivo de los ítems que conforman la prueba de matemáticas de tercer grado a través del análisis de contenido del razonamiento al momento de responder y su correspondencia con el nivel cognitivo indicado en la tabla de especificaciones de la prueba, como evidencia de validez del nivel cognitivo.

C. Variables

La variable del estudio consiste en la fuente de validez de la prueba de matemática de tercer grado:

- Validez predictiva
- Validez de constructo
- Validez de contenido (del nivel cognitivo)

D. Enfoque de investigación

Para cumplir con los objetivos propuestos, se pretende analizar la validez predictiva, de constructo y del nivel cognitivo de ítems de la prueba de matemáticas de tercer grado,

bajo un enfoque mixto de investigación. Para determinar si existe correlación entre los resultados obtenidos en la prueba de matemáticas de tercer grado y en la prueba de matemáticas de sexto grado y para identificar los factores que conforman la estructura de la prueba, se realizan los análisis correspondientes desde el enfoque cuantitativo. Mientras que para establecer la correspondencia entre el nivel de razonamiento al momento de responder a cada ítem de la prueba de matemáticas, y el nivel cognitivo propuesto en la tabla de especificaciones basado en la taxonomía de Marzano, el análisis se realiza desde el enfoque cualitativo.

E. Tipo de investigación

Basado en el enfoque cuantitativo, el tipo de investigación es no experimental de diseño transversal correlacional. La investigación no experimental se refiere a estudios que se realizan sin la manipulación deliberada de variables y en los que sólo se observan los fenómenos en su ambiente natural para después analizarlos. Los diseños correlacionales describen relaciones entre dos o más categorías, conceptos o variables en un momento determinado (Hernández, Fernández y Baptista, 2010).

Bajo el enfoque cualitativo, el tipo de investigación corresponde a un diseño fenomenológico, el cual se enfoca en las experiencias individuales subjetivas de los participantes, en donde se pretende reconocer las percepciones de las personas y el significado de un fenómeno o experiencia (Hernández *et al.*, 2010).

F. Hipótesis

H0: No existe correlación estadísticamente significativa a nivel 0.05 entre los resultados de la prueba de matemáticas de tercer grado y los resultados obtenidos por los mismos niños en la prueba de matemáticas de sexto grado.

H1: Existe correlación estadísticamente significativa a nivel 0.05 entre los resultados de la prueba de matemáticas de tercer grado y los resultados obtenidos por los mismos niños en la prueba de matemáticas de sexto grado.

G. Supuestos de investigación

Desde la perspectiva cualitativa, se espera que el análisis del discurso de los alumnos al enfrentarse a los ítems de la prueba de matemáticas de tercer grado, coincida con la clasificación que el experto redactor de ítems asignó a cada ítem de la prueba.

H. Sujetos de investigación

Los sujetos de estudio abarcan 237 niños y niñas de tercer grado, 112 de sexo femenino y 125 de sexo masculino, comprendidos entre las edades de 8 y 14 años, con una media de 9.62 años y desviación típica de 1.22 años. Pertenecientes a establecimientos educativos de San Miguel Escobar, San Miguel Milpas Altas, San Rafael El Arado y Santa Catarina Bobadilla. Todos ellos fueron evaluados en la prueba de Matemática de tercer grado en el año 2011. Asimismo un grupo de 126 de estos niños que en el año 2014 cursaron sexto grado y que fueron evaluados con la prueba de Matemática de sexto grado que aplica el Centro de Investigaciones Educativas de la Universidad del Valle de Guatemala. De igual forma, 21 niños (9 niñas y 12 niños), que finalizaron tercer grado en el año 2012 y que fueron evaluados con la prueba de matemáticas por bloques de ítems; es decir, no todos los niños respondieron a todos los ítems, cuyas respuestas orales fueron grabadas y posteriormente transcritas para el respectivo análisis de contenido.

I. Instrumento

La prueba de matemática de tercer grado creada por el Centro de Investigaciones Educativas de la Universidad del Valle, está conformada por 40 ítems de selección múltiple con cuatro opciones de respuesta cada uno, tanto en la forma A como en la forma B, que miden *Operaciones* (suma, resta, multiplicación y división): ítems 1 al 12, *Conceptos aritméticos* (numeración, geometría, medidas, fracciones, gráficas y comprensión de operaciones): ítems 13 al 28 y *Resolución de problemas* (suma, resta, multiplicación, división, varias operaciones y lógica): ítems 29 al 40, de acuerdo a la tabla de especificaciones elaborada por el Centro de Investigaciones Educativas (ver Apéndices).

J. Estadísticos

Para comprobar la correlación entre los resultados obtenidos en la prueba de matemáticas de tercer grado y en la prueba de matemáticas de sexto grado, se utiliza el coeficiente *r de Pearson* y un análisis de *regresión lineal simple*.

Para corroborar la estructura de la prueba de matemáticas de tercer grado se realiza un *análisis factorial confirmatorio de componentes principales con rotación varimax*, sobre los cuatro niveles de procesamiento cognitivo (conocimiento, comprensión, análisis y utilización) de acuerdo a la taxonomía de Marzano, a través del programa SPSS versión 20.

K. Procedimiento

- Los niños fueron evaluados por el CIE para una consultoría de evaluación, en la prueba de matemática de tercer grado en el año 2011 y en la prueba de matemática de sexto primaria en el año 2014.
- Se recolectaron las bases de datos proporcionadas por el Centro de Investigaciones Educativas para realizar los análisis correlacional y factorial.
- Los resultados de los análisis correlacional y factorial se presentan en tablas y gráficas para una mejor comprensión de los mismos.
- Para el análisis de contenido, las respuestas orales de los niños fueron grabadas y posteriormente transcritas por estudiantes del curso Investigación Cualitativa de la maestría en educación de la Universidad del Valle en el año 2012.
- El análisis cualitativo presenta el contenido intencional o propuesto (contenido de área) para cada ítem, basado en la tabla de especificaciones elaborada por el Centro de Investigaciones Educativas.
- Enseguida se presenta el contenido anticipado (procesos cognitivos), basado en la taxonomía de Marzano, y señalado mediante el juicio de expertos, por parte de estudiantes del curso Investigación Cualitativa de las Maestrías en Educación de la Universidad del Valle e incluye también el juicio de la responsable de esta investigación.

- Por último se presenta el contenido actual (estrategias de respuestas de los examinados), basado en los protocolos de “pensamiento en voz alta” de los niños de tercero primaria que respondieron a la prueba (ver apéndices).

L. Alcances y límites

El alcance de la presente investigación pretende determinar distintas fuentes de validez de la prueba de matemática de tercer grado que aplica el Centro de Investigaciones Educativas de la Universidad del Valle de Guatemala. Por lo tanto, los resultados que se obtengan a partir de los análisis que se realicen, se circunscriben únicamente al escenario propuesto en el estudio, es decir, no pueden ser generalizados a toda la población guatemalteca y tampoco a los resultados de otras pruebas del área de matemáticas, o distintas fuentes de validez que no se incluyen en el estudio, como sería la validez concurrente, la cual indica el grado en el que las puntuaciones de una prueba pueden servir para estimar la posición actual de un individuo frente a un criterio casi al mismo tiempo (Cohen y Swerdlik, 2006).

V. RESULTADOS

A continuación se presentan las tablas y gráficos obtenidos al realizar los análisis de correlación, regresión y factorial, como evidencia de la validez predictiva y de constructo de los resultados de la prueba de matemática de tercer grado. Asimismo, se incluyen los contenidos alineados y no alineados, como evidencia de la validez cognitiva de los ítems de la prueba.

A. Validez predictiva

Tal como mencionan Kerligner y Lee (2002), la validez predictiva involucra el uso de desempeños de criterio futuros y está determinada por el grado de correspondencia entre las medidas implicadas. Según indica Nunnally (1987), la validez de los instrumentos predictores se determina por medio del análisis correlacional.

En cuanto a la correlación de los resultados de la prueba de matemática de tercer grado y la de sexto grado, se buscó la coincidencia de las puntuaciones de respuestas correctas obtenidas por los mismos sujetos. De tal cuenta que resultaron 126 niños y niñas de los 237 que fueron evaluados al final del año 2011 en tercer grado y que permanecieron en las mismas escuelas y fueron evaluados al final del año 2014 en sexto grado.

Por lo tanto, se conformó una base de datos con los niños que fueron evaluados en la prueba de Matemática de tercer grado en el año 2011, quedando de esta forma un grupo de 126 niños que en el año 2014 se encontraban en sexto grado y que también fueron evaluados con la prueba de Matemática de sexto primaria que aplica el CIE.

En la Tabla 1 se observa el promedio de respuestas correctas obtenido por los niños y niñas en la prueba de matemáticas de tercer grado y de sexto grado, así como la desviación típica en cada una de ellas. Dicho promedio se encuentra sobre un total de 40 ítems evaluados tanto en la prueba de tercero como en la de sexto grado.

Tabla 1. Estadísticos descriptivos

	Media	Desviación típica	N
Respuestas correctas Tercero	20.4921	7.40675	126
Respuestas correctas Sexto	23.8333	6.73112	126

Al realizar el análisis de correlación entre los resultados de ambas pruebas, se obtuvo un coeficiente de correlación *r de Pearson* de 0.646, el cual es estadísticamente significativo a nivel 0.01, tal como se muestra en la Tabla 2.

Tabla 2. Correlaciones

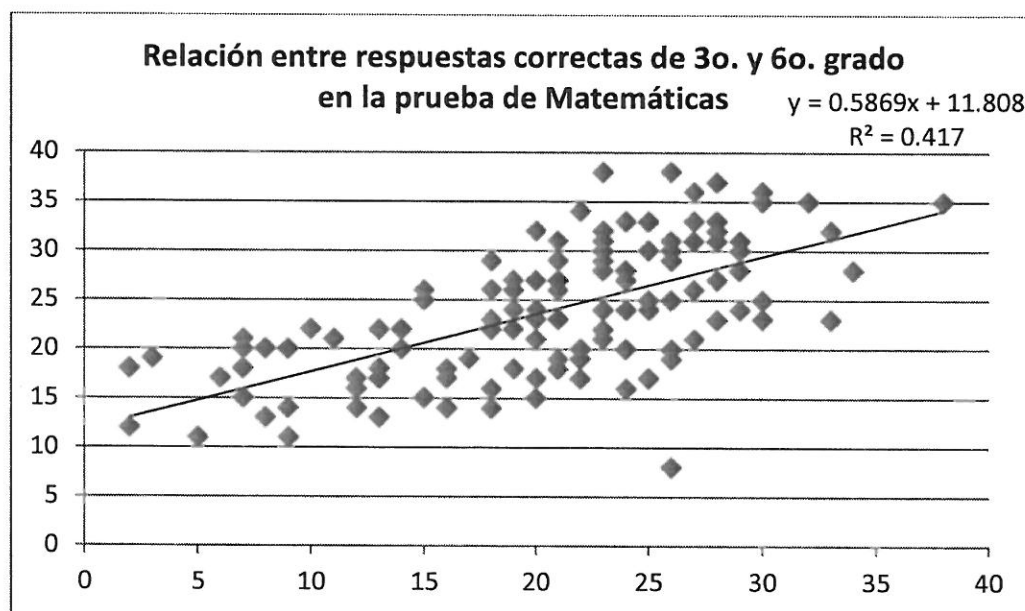
		Respuestas correctas Tercero	Respuestas correctas Sexto
Respuestas correctas Tercero	Correlación de Pearson	1	.646**
	Sig. (bilateral)		.000
	N	126	126
Respuestas correctas Sexto	Correlación de Pearson	.646**	1
	Sig. (bilateral)	.000	
	N	126	126

** . La correlación es significativa al nivel 0,01 (bilateral).

Asimismo se realizó un análisis de regresión lineal simple, ya que se trata de dos variables, el cual consiste en obtener una mejor estimación de Y , dado cualquier valor de X , y se aplica cuando existe una relación lineal entre X y Y en un diagrama de dispersión, que constituye una representación gráfica de los datos. La fórmula de la ecuación lineal (o fórmula de la recta de regresión) es $\hat{Y} = a + bX$ (Ritchey, 2008).

Al observar el Gráfico 1 se puede notar que el eje X corresponde a las respuestas correctas obtenidas en la prueba de matemáticas de tercer grado y en el eje Y se representan las puntuaciones de respuestas correctas en la prueba de matemáticas de sexto grado. El coeficiente de determinación (R^2) es 0.417, lo cual indica que los resultados de la prueba de matemáticas de tercer grado explican un 41.7% de varianza de la prueba de matemáticas de sexto grado.

Gráfico 1. Gráfico de correlación



Se observa, tanto en el Gráfico 1 como en la Tabla 3 que $a=11.808$ representa la intersección con Y , es decir el punto en el que la recta de regresión interseca al eje Y cuando $X=0$ y recibe el nombre de *constante* de la ecuación. Y $b=0.587$ se refiere a la *pendiente* de la recta de regresión, denominada *coeficiente de regresión*, es decir el efecto sobre Y de un cambio de una unidad en X (Ritchey, 2008).

Tabla 3. Regresión

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.646 ^a	.417	.412	5.16019

a. Variables predictoras: (Constante), Respuestas correctas Tercero

Tabla 3 (continuación). Coeficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
(Constante)	11.808	1.357		8.700	.000
1 Respuestas correctas Tercero	.587	.062	.646	9.418	.000

a. Variable dependiente: Respuestas correctas Sexto

Con lo anterior se llega a determinar que los resultados alcanzados en la prueba de matemáticas de tercer grado son responsables en un 41.7% de los resultados que los mismos estudiantes obtienen en la prueba de matemáticas de sexto grado. Es decir, la puntuación en la prueba de sexto, depende en un 42% (valor aproximado) de la calificación conseguida por el mismo estudiante en la prueba de tercer grado. Estos resultados apoyan la evidencia de validez predictiva de las puntuaciones de la prueba de tercer grado, por el nivel de correspondencia entre las medidas implicadas y que permite diferenciar a estudiantes que obtiene punteos altos y los que obtienen punteos bajos en un criterio futuro.

B. Validez de constructo

La evidencia convergente y discriminativa de la validez de constructo puede obtenerse por medio del uso del análisis factorial. El análisis factorial se utiliza para identificar factores o variables específicas que de manera particular son atributos, características o dimensiones en las que es posible diferir. En la investigación psicométrica, es empleado con frecuencia como un método de reducción de datos en el que son analizados varios conjuntos de puntuaciones y las correlaciones entre ellas. El propósito del análisis factorial puede ser identificar el factor o factores en común entre las puntuaciones de la prueba, en subescalas dentro de una prueba particular, o los factores en común entre puntuaciones obtenidas en una serie de pruebas (Cohen y Swerdlik, 2006).

Según describe De la Fuente (2011), el análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Los grupos homogéneos se forman con las variables que

correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

El análisis factorial es, por tanto, una técnica de reducción de las dimensiones de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos (De la Fuente, 2011).

El análisis factorial se realiza ya sea en una base exploratoria o en una base confirmatoria. El análisis factorial exploratorio implica la estimación o extracción de factores, la decisión de cuántos factores conservar y la rotación de estos a una orientación interpretable. Mientras que el análisis factorial confirmatorio se plantea la hipótesis de una estructura factorial en forma explícita y se prueba su ajuste con la estructura de la covarianza observada en las variables medidas (Floyd y Widaman, citados en Cohen y Swerdlik, 2006).

Al cargar un factor en una prueba se transmite información del grado en que el factor determina la puntuación o puntuaciones de la prueba. De esta manera, las cargas factoriales altas en una prueba proporcionan evidencia convergente de la validez de constructo; las cargas moderadas a bajas proporcionan evidencia discriminativa de la validez de constructo (Cohen y Swerdlik, 2006).

Según estos mismos autores, la denominación de factores que surgen de un análisis factorial tiene más que ver con el conocimiento, el juicio y la abstracción verbal y no hay reglas específicas para designar factores, sino solamente se ejerce un juicio concerniente a qué nombre del factor común comunica mejor su significado.

El método de extracción más usual es el de Análisis de Componentes Principales. Asimismo se observan las *estimaciones de la varianza* que cada ítem tiene en común con los demás y que se denominan Comunalidades, en donde los ítems 34 y 37 son los que comparten menor varianza con los demás ítems, tal como se observa en la Tabla 4.

Tabla 4.
Comunalidades

	Inicial	Extracción
ítem 1 en el post-test	1.00	0.735
ítem 2 en el post-test	1.00	0.620
ítem 3 en el post-test	1.00	0.486
ítem 4 en el post-test	1.00	0.832
ítem 5 en el post-test	1.00	0.841
ítem 6 en el post-test	1.00	0.353
ítem 7 en el post-test	1.00	0.767
ítem 8 en el post-test	1.00	0.407
ítem 9 en el post-test	1.00	0.503
ítem 10 en el post-test	1.00	0.223
ítem 11 en el post-test	1.00	0.347
ítem 12 en el post-test	1.00	0.470
ítem 13 en el post-test	1.00	0.356
ítem 14 en el post-test	1.00	0.811
ítem 15 en el post-test	1.00	0.495
ítem 16 en el post-test	1.00	0.167
ítem 17 en el post-test	1.00	0.498
ítem 18 en el post-test	1.00	0.674
ítem 19 en el post-test	1.00	0.769
ítem 20 en el post-test	1.00	0.505
ítem 21 en el post-test	1.00	0.213
ítem 22 en el post-test	1.00	0.606
ítem 23 en el post-test	1.00	0.531
ítem 24 en el post-test	1.00	0.322
ítem 25 en el post-test	1.00	0.236
ítem 26 en el post-test	1.00	0.608
ítem 27 en el post-test	1.00	0.156
ítem 28 en el post-test	1.00	0.554
ítem 29 en el post-test	1.00	0.503
ítem 30 en el post-test	1.00	0.700
ítem 31 en el post-test	1.00	0.476
ítem 32 en el post-test	1.00	0.391
ítem 33 en el post-test	1.00	0.282
ítem 34 en el post-test	1.00	0.064
ítem 35 en el post-test	1.00	0.543
ítem 36 en el post-test	1.00	0.179
ítem 37 en el post-test	1.00	0.080
ítem 38 en el post-test	1.00	0.272
ítem 39 en el post-test	1.00	0.370
ítem 40 en el post-test	1.00	0.294

Método de extracción: Análisis de Componentes principales.

El método de extracción de componentes principales permite elegir todos los factores cuyos autovalores excedan un valor especificado (por defecto 1). Se observa en la Tabla 5 que al realizar la extracción de cuatro factores, el porcentaje de la varianza explicada es 45.6%, el cual corresponde al porcentaje acumulado de los primeros cuatro componentes extraídos con autovalores iniciales más altos. Generalmente, un conjunto pequeño de factores, los primeros, contienen casi toda la información y el resto de factores suelen contribuir relativamente poco (De la Fuente, 2011). Lo que se busca es que entre todos los factores (o nuevas medidas) se explique la mayor proporción de varianza posible. No suelen extraerse todos los factores posibles, sino solamente los que explican al menos más varianza que cualquiera de las variables tomada aisladamente; aunque un menor número de factores o componentes explican menos varianza, la estructura es más clara e interpretable (Morales, 2013).

Tabla 5.
Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6.517	16.293	16.293	6.517	16.293	16.293
2	4.765	11.912	28.205	4.765	11.912	28.205
3	3.958	9.894	38.099	3.958	9.894	38.099
4	3.001	7.503	45.602	3.001	7.503	45.602
5	2.801	7.004	52.605			
6	2.513	6.281	58.886			
7	2.440	6.099	64.985			
8	2.118	5.294	70.280			
9	1.661	4.152	74.432			
10	1.548	3.869	78.301			
11	1.320	3.300	81.601			
12	1.242	3.104	84.705			
13	1.181	2.954	87.659			
14	0.984	2.459	90.118			
15	0.770	1.925	92.043			
16	0.658	1.645	93.688			
17	0.598	1.494	95.182			
18	0.504	1.261	96.443			
19	0.374	0.935	97.377			
20	0.323	0.806	98.184			
21	0.287	0.717	98.900			
22	0.247	0.618	99.519			
23	0.116	0.291	99.810			
24	0.076	0.190	100			
25	7.59E-16	1.90E-15	100			
26	5.13E-16	1.28E-15	100			
27	4.20E-16	1.05E-15	100			
28	3.81E-16	9.53E-16	100			
29	2.43E-16	6.06E-16	100			
30	1.74E-16	4.35E-16	100			
31	8.05E-17	2.01E-16	100			
32	2.78E-17	6.94E-17	100			
33	-4.64E-19	-1.16E-18	100			
34	-1.07E-16	-2.67E-16	100			
35	-1.82E-16	-4.55E-16	100			
36	-3.19E-16	-7.97E-16	100			
37	-3.75E-16	-9.38E-16	100			
38	-4.55E-16	-1.14E-15	100			
39	-5.37E-16	-1.34E-15	100			
40	-8.53E-16	-2.13E-15	100			

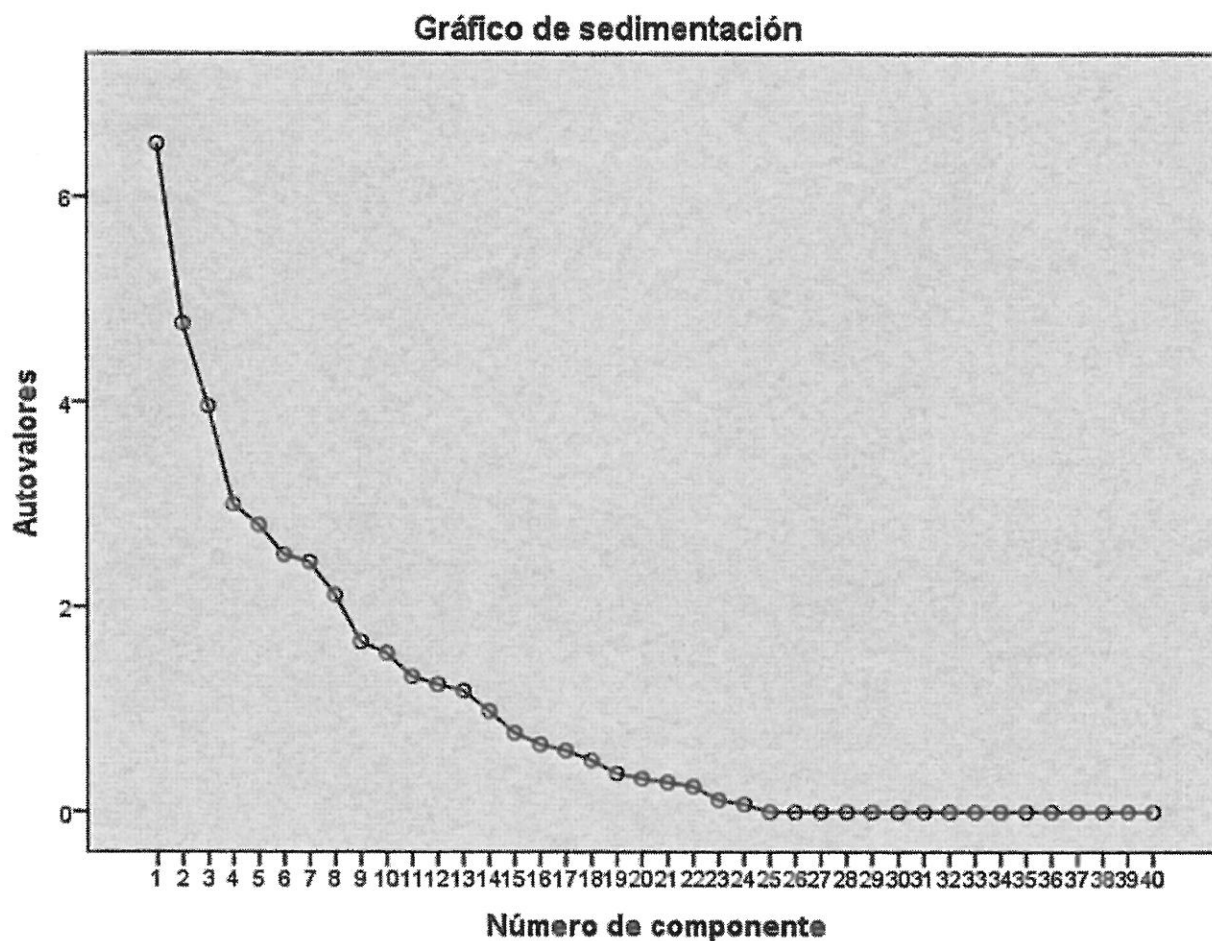
Tabla 5. (continuación)

Componente	Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado
1	6.392	15.980	15.980
2	4.648	11.620	27.600
3	3.970	9.924	37.524
4	3.231	8.078	45.602
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			

Método de extracción: Análisis de Componentes principales

El Gráfico 2 de sedimentación justifica la selección de los factores, por lo que se busca el punto de inflexión en la gráfica. Y del mismo se deriva que es posible extraer cuatro factores. Según Morales (2013) el punto de corte para establecer el número de factores que se van a rotar se sitúa en el punto de inflexión de la línea descendente que va uniendo los diversos eigenvalues.

Gráfico 2.



El método de rotación Varimax es el más frecuentemente utilizado y está orientado a la maximización de la varianza de los factores y con él se reduce el número de factores que son necesarios para explicar una variable, en este caso la variable se refiere a los niveles de procesamiento cognitivo (conocimiento, comprensión, análisis y utilización) según la taxonomía de Marzano.

En la Tabla 6, sobre la matriz factorial de componentes rotados, aparecen juntos los ítems con cargas altas para el mismo factor, es decir las variables que tienen correlaciones altas en ese factor. Se eliminan las *cargas factoriales bajas* (debajo de 0.30), debido a que el investigador decide a partir de qué valor deben eliminarse las cargas factoriales y de este modo se suprime información redundante (De la Fuente, 2011).

Para que un ítem se considere que pertenece a un factor, o dicho con más propiedad, que lo define o explica suficientemente, debe tener en él una correlación razonablemente alta (por lo menos .30, o más) y no tenerlas mayores en otro factor (Morales, 2013).

Asimismo, de acuerdo a Nunnally (citado en Morales, 2013) para reducir el número de factores de manera no arbitraria se pueden encontrar varias orientaciones (como eliminar los factores en los que ninguna variable tiene un peso superior a .30).

Desde el punto de vista de la relevancia suele considerarse un valor en torno a .30 como mínimo (explicaría aproximadamente el 10% de la varianza); en torno a .40 ya es más relevante, y valores en torno a .50 son ya de clara relevancia práctica y definen bien el factor (Morales, 2013).

En la Tabla 6 asimismo se observa que algunos ítems muestran una carga considerable en varios factores a la vez. Mientras que los ítems 17, 25 y 40 no poseen cargas factoriales mayores a 0.30 en ninguno de los factores.

Tabla 6.
Matriz de componentes rotados^a

	Componente			
	1	2	3	4
ítem 14 en el post-test	0.868			
ítem 4 en el post-test	0.865			
ítem 19 en el post-test	-0.790		0.331	
ítem 1 en el post-test	-0.771			
ítem 30 en el post-test	-0.737		-0.303	0.329
ítem 9 en el post-test	0.654			
ítem 3 en el post-test	0.633	0.445		
ítem 26 en el post-test	0.623			-0.309
ítem 33 en el post-test	0.520		-0.325	
ítem 24 en el post-test	-0.472			
ítem 36 en el post-test	-0.461			
ítem 23 en el post-test	-0.395			
ítem 27 en el post-test		0.785		
ítem 10 en el post-test		-0.770		
ítem 18 en el post-test		0.635		
ítem 2 en el post-test		-0.569		
ítem 20 en el post-test		-0.568		
ítem 15 en el post-test		0.559		-0.305
ítem 8 en el post-test	-0.332	0.521		
ítem 11 en el post-test	0.394	0.459		
ítem 12 en el post-test	-0.357	0.445		
ítem 32 en el post-test		0.405	0.395	-0.379
ítem 29 en el post-test		0.376		-0.331
ítem 17 en el post-test				
ítem 38 en el post-test			-0.860	
ítem 21 en el post-test			0.771	-0.395
ítem 37 en el post-test			0.597	0.495
ítem 34 en el post-test			0.572	
ítem 7 en el post-test			0.526	
ítem 5 en el post-test			0.526	-0.394
ítem 28 en el post-test			0.450	0.327
ítem 6 en el post-test			0.377	
ítem 13 en el post-test				-0.657
ítem 31 en el post-test		0.434		0.565
ítem 35 en el post-test				0.533
ítem 22 en el post-test				0.424
ítem 39 en el post-test				0.407
ítem 16 en el post-test		-0.311		0.323
ítem 25 en el post-test				
ítem 40 en el post-test				

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 6 iteraciones.

Una forma de llevar a cabo la selección de variables (ítems) es estudiar la *matriz de correlaciones* de las variables (ítems) con los factores, seleccionando como representante de cada factor la variable (ítem) con la correlación más elevada en éste, que sea más fácil de medir y que tenga más sentido desde un punto de vista teórico. En cualquier caso, conviene elegir las variables (ítems) de forma que una misma variable o ítem no se utilice para medir dos factores distintos (De la Fuente, 2011).

El análisis factorial indica cómo tienden a agruparse los ítems o variables. Examinando el contenido conceptual de los ítems que pertenecen al mismo factor se puede comprender qué factores [o constructos] subyacentes explican las correlaciones entre los ítems. Los factores equivalen a constructos hipotéticos o conceptos subyacentes o latentes (no observables directamente) deducidos de las correlaciones entre las variables. (Morales, 2013).

De manera que después de las rotaciones viene una fase de reflexión o análisis conceptual. Los factores indican cómo tienden a agruparse los ítems o variables. Y se trata de encontrar significado a los factores: ¿qué tienen en común los ítems que mejor definen cada factor?; ¿cómo se puede denominarlos ('bautizarlos')?; ¿cómo aclaran el constructo subyacente a todos ellos?, etc. Se les suele denominar con una expresión que refleje el significado común a todos los ítems que expresan cada factor (Morales, 2013).

Es necesario además hacer un análisis conceptual y ver qué tienen en común las formulaciones de los ítems que definen el mismo factor, sobre todo las de los ítems que tienen los pesos mayores, que son los que mejor lo definen. Y después de este análisis hay que dar un nombre a estos factores (Morales, 2013).

De esta manera, en la Tabla 7 los ítems que se agrupan en el primer factor son: ítems 1 y 3 que corresponden a sumas de dos sumandos (de tres dígitos "sin llevar" y de cuatro dígitos "llevando", respectivamente); el ítem 4 (resta de tres dígitos en el sustraendo "sin prestar"); ítem 9 (multiplicación de tres cifras en el multiplicando y una cifra en el multiplicador "llevando"); ítem 14 (reconocimiento de una figura geométrica: cuadrado o triángulo); ítem 19 (reconocer el conjunto que tiene sombreadas la mitad de las figuras);

ítem 23 (saber qué hora tiene el reloj y cuánto falta para la siguiente hora); ítem 24 (saber en qué se mide la distancia de una carretera o la altura de un árbol); ítem 26 (saber qué número debe sumar a seis o a siete para que dé diez); ítem 30 (resolver un problema de suma de tres dígitos en los sumandos “sin llevar”); ítem 33 (resolver un problema de suma y multiplicación de un dígito) y el ítem 36 (resolver un problema de multiplicación de un dígito por 10, 20 ó 30 y suma de dos sumandos de dos cifras terminadas en cero). Respecto a estos ítems se podría decir que tienen en común el *conocimiento* básico sobre operaciones de suma, resta, multiplicación, geometría, fracciones y medidas, ya que requieren no solo cierto nivel de reconocimiento, sino además la producción de información almacenada (Marzano y Kendall, 2008).

Asimismo en la Tabla 7, los ítems que se agrupan en el segundo factor son: ítem 2 (suma de tres sumandos de tres cifras “sin llevar”); ítem 8 (multiplicación de dos cifras en el multiplicando y una cifra en el multiplicador “llevando”, pero los números del multiplicando son “siete, ocho o nueve”); ítems 10, 11 (división de dos dígitos en el dividendo y un dígito en el divisor); ítem 12 (división de tres dígitos en el dividendo y un dígito en el divisor); ítem 15 (determinar el número que aparece en dos de tres figuras geométricas superpuestas); ítem 18 (identificar la figura que está dividida en tercios); ítem 20 (determinar el número que va antes de otro, contando de uno en uno); ítem 27 (encontrar el minuendo de una resta de un dígito, dados el sustraendo y la diferencia); ítem 29 (resolver el primer problema de la serie de resolución de problemas, una suma de dos sumandos en donde uno de ellos es cien); ítem 32 (resolver un problema de división de dos dígitos en el dividendo y en el divisor, terminados en cero). Estos ítems conllevan el conocimiento básico de operaciones no solo de suma y multiplicación, sino además de división, así como se podría deducir un nivel de *comprensión* de geometría, fracciones y numeración, puesto que envuelven extraer el conocimiento hacia características clave organizadas en una forma generalizada como información y eventos generales (Marzano y Kendall, 2008).

De igual forma en la Tabla 7 los ítems que se agrupan en el tercer factor son: ítem 5 (resta de tres dígitos en el minuendo y dos dígitos en el sustraendo “sin prestar”); ítem 6 (resta de cuatro dígitos en el minuendo y en el sustraendo “prestando”); ítem 7

(multiplicación de una cantidad de tres dígitos por dos); ítem 21 (indicar el número que va después de otro, contando de dos en dos); ítem 28 (indicar de qué otra manera se puede decir una multiplicación); ítem 34 (resolver un problema de resta de tres dígitos); ítem 37 (resolver un problema de división de dos dígitos en el dividendo y en el divisor e indicar el residuo); ítem 38 (resolver un problema de multiplicación y con el resultado realizar una adición). Estos ítems, adicional al conocimiento básico de operaciones como resta y multiplicación, se podría decir que involucran un nivel de *utilización* de este conocimiento, proceso que es utilizado por el individuo cuando desea completar una tarea específica como seleccionar entre dos o más alternativas, identificar los criterios que se usan al juzgar el valor de cada alternativa y por último, seleccionar entre las alternativas basadas en la aplicación del criterio (Marzano y Kendall, 2008).

Y finalmente, también en la Tabla 7, los ítems que se agrupan en el cuarto factor son: ítem 13 (responder cuántos lados tiene un rombo); ítem 16 (indicar cuánto mide un rectángulo en sus cuatro lados); ítem 22 (con base en un cuadro, señalar el día que se vendió más pan/tortillas); ítem 31 (resolver un problema de resta de tres dígitos en el minuendo y dos dígitos en el sustraendo “sin prestar”); ítem 35 (resolver un problema de multiplicación de dos dígitos por un dígito); ítem 39 (resolver un problema sobre cómo puede averiguar cuánto le sobró de algo). Estos ítems se podría decir que incluyen el *análisis* tanto de figuras geométricas como en la resolución de problemas, lo cual implica generar nuevas aplicaciones de una generalización o principio conocidos (Marzano y Kendall, 2008).

Tabla 7. Factores

Ítems	Factor 1 Conocimiento	Factor 2 Comprensión	Factor 3 Utilización	Factor 4 Análisis
Ítem1	-0.771			
Ítem 3	0.633	0.445		
Ítem 4	0.865			
Ítem 9	0.654			
Ítem 14	0.868			
Ítem 19	-0.79		0.331	
Ítem 23	-0.395			
Ítem 24	-0.472			
Ítem 26	0.623			-0.309
Ítem 30	-0.737		-0.303	0.329
Ítem 33	0.52		-0.325	
Ítem 36	-0.461			
Ítem 2		-0.569		
Ítem 8	-0.332	0.521		
Ítem 10		-0.77		
Ítem 11	0.394	0.459		
Ítem 12	-0.357	0.445		
Ítem 15		0.559		-0.305
Ítem 18		0.635		
Ítem 20		-0.568		
Ítem 27		0.785		
Ítem 29		0.376		-0.331
Ítem 32		0.405	0.395	-0.379
Ítem 5			0.526	-0.394
Ítem 6			0.377	
Ítem 7			0.526	
Ítem 21			0.771	-0.395
Ítem 28			0.45	0.327
Ítem 34			0.572	
Ítem 37			0.597	0.495
Ítem 38			-0.86	
Ítem 13				-0.657
Ítem 16		-0.311		0.323
Ítem 22				0.424
Ítem 31		0.434		0.565
Ítem 35				0.533
Ítem 39				0.407
% de varianza explicada	15.98%	11.62%	9.92%	8.08%

C. Validez del nivel cognitivo de los ítems

Para realizar este análisis se contó con 21 niños (9 niñas y 12 niños), que finalizaron tercer grado en el año 2012 y que fueron evaluados con la prueba de matemáticas por bloques de ítems; es decir, no todos los niños respondieron a todos los ítems. Fueron analizados por estudiantes del curso Investigación Cualitativa de las diferentes maestrías en educación de la Universidad del Valle de Guatemala. Las respuestas orales de los niños fueron grabadas y posteriormente transcritas para el respectivo análisis de contenido.

Cada uno de los ítems de la prueba y en algunos de ellos en forma A y en forma B, fueron analizados en cuanto al contenido intencional o propuesto (contenido de área), basado en la tabla de especificaciones elaborada por el Centro de Investigaciones Educativas. Asimismo el contenido anticipado (procesos cognitivos), basado en la taxonomía de Marzano, y señalado mediante el juicio de expertos, por parte de profesores y en este caso, también estudiantes del curso Investigación Cualitativa de las maestrías en Educación de la Universidad del Valle, así como el juicio de la responsable del presente trabajo de tesis. Por último el contenido actual (estrategias de respuestas de los examinados), basado en los protocolos de “pensamiento en voz alta” de los niños que respondieron a la prueba.

En las tablas 8, 9 y 10 se muestra si los ítems se encuentran “*alineados, parcialmente alineados o no alineados*”, de acuerdo al siguiente criterio: si coinciden el **contenido intencional** (propuesto) y el **contenido anticipado** (basado en los juicios de expertos sobre los protocolos de “pensamiento en voz alta” de los niños) en su totalidad en cuanto al proceso cognitivo principal (conocimiento, comprensión, análisis y utilización) que refleja el ítem, entonces se dice que el ítem está “*alineado*”; si difieren el **contenido intencional** y el **contenido anticipado** en alguno de los procesos cognitivos (conocimiento, comprensión, análisis y utilización) que refleja el ítem, se dice que el ítem está “*parcialmente alineado*” y si **ambos contenidos, intencional y anticipado**, difieren en su totalidad, se dice que el ítem está “*no alineado*”.

Tabla 8. Ítems alineados

Ítems	Contenido intencional/propuesto (contenido de área) →Objetivos educativos/taxonomía	Contenido anticipado (procesos cognitivos) →Expertos en contenido/profesores	Resultado
Ítem 1 "A"	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar)	Alineado
Ítem 3 "A"	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar)	Alineado
Ítem 5 "A y B"	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar)	Alineado
Ítem 7 "A"	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar)	Alineado
Ítem 10 "A"	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar)	Alineado
Ítem 12 "A"	Conocimiento/recuerdo	Conocimiento/recuerdo	Alineado
Ítem 13 "A y B"	Comprensión	Comprensión (representación)	Alineado
Ítem 14 "A"	Comprensión	Comprensión (síntesis) Comprensión (representación)	Alineado
Ítem 15 "A"	Análisis	Análisis (relación)	Alineado
Ítem 15 "B"	Análisis	Análisis (clasificación) Análisis (relación)	Alineado
Ítem 16 "A"	Análisis	Análisis (especificación)	Alineado
Ítem 19 "A"	Comprensión	Comprensión (síntesis) Comprensión (representación)	Alineado
Ítem 22 "A"	Análisis	Análisis (clasificación) Análisis (relación)	Alineado
Ítem 22 "B"	Análisis	Análisis (clasificación) Análisis (relación)	Alineado
Ítem 23 "A"	Comprensión	Comprensión (síntesis)	Alineado
Ítem 26 "B"	Análisis	Análisis Análisis (clasificación)	Alineado
Ítem 29 "A"	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 30 "A"	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 31 "A"	Utilización	Utilización (resolución de problemas)	Alineado

Tabla 8. (continuación)			
Ítems	Contenido intencional/propuesto (contenido de área) →Objetivos educativos/taxonomía	Contenido anticipado (procesos cognitivos) →Expertos en contenido/profesores	Resultado
Ítem 31 “B”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 32 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 33 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 34 “B”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 35 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 37 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 38 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 38 “B”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 39 “A”	Utilización	Utilización (resolución de problemas)	Alineado
Ítem 40 “A”	Utilización	Utilización (toma de decisiones) Utilización (resolución de problemas)	Alineado
Ítem 40 “B”	Utilización	Utilización (toma de decisiones)	Alineado

De la Tabla 8 se deriva que treinta ítems del total de los cincuenta ítems analizados están “alineados”, es decir representan un 60% de “alineación” de la prueba.

Estos ítems se refieren, del 1 al 12 a **operaciones** y los que se encuentran “alineados” representan un 50% de “alineación” en este contenido (ítems 1, 3, 5, 7, 10 y 12). Los ítems “alineados” que corresponden a **conceptos aritméticos**, que van del 13 al 28, representan un 50% de “alineación” en dicho contenido (ítems 13, 14, 15, 16, 19, 22, 23 y 26). Y de los ítems 29 al 40 que corresponden a **resolución de problemas**, los que se encuentran “alineados” representan un 92% de “alineación” en este contenido (ítems 29, 30, 31, 32, 33, 34, 35, 37, 38, 39 y 40).

Tabla 9. Ítems parcialmente alineados

Ítems	Contenido intencional/propuesto (contenido de área) →Objetivos educativos/taxonomía	Contenido anticipado (procesos cognitivos) →Expertos en contenido/profesores	Resultado
Ítem 2 “B”	Conocimiento/recuerdo	Conocimiento/recuerdo (ejecutar) Utilización	Parcialmente alineado
Ítem 4 “B”	Conocimiento/recuerdo	Conocimiento/recuerdo Utilización	Parcialmente alineado
Ítem 11 “B”	Conocimiento/recuerdo	Conocimiento/recuerdo Utilización	Parcialmente alineado
Ítem 14 “B”	Comprensión	Conocimiento/recuerdo (ejecutar) Comprensión (síntesis)	Parcialmente alineado
Ítem 16 “B”	Análisis	Análisis (especificación) Utilización	Parcialmente alineado
Ítem 17 “B”	Conocimiento/recuerdo	Conocimiento/recuerdo (nombrar) Utilización	Parcialmente alineado
Ítem 29 “B”	Utilización	Comprensión (síntesis) Utilización (resolución de problemas)	Parcialmente alineado
Ítem 30 “B”	Utilización	Análisis (relación) Utilización (resolución de problemas)	Parcialmente alineado
Ítem 36 “B”	Utilización	Análisis (especificación) Utilización (resolución de problemas)	Parcialmente alineado
Ítem 37 “B”	Utilización	Análisis (especificación) Utilización (resolución de problemas)	Parcialmente alineado

En la Tabla 9 se observa que diez ítems del total de los cincuenta ítems analizados, representan un 20% de “alineación parcial” de la prueba. De ellos, un 25% (ítems 2, 4 y 11) corresponden al contenido de **operaciones**, que van del 1 al 12 (suma, resta, multiplicación y división); un 19% (ítems 14, 16 y 17) se refieren al contenido de **conceptos aritméticos**, que van del 13 al 28 (numeración, geometría, medidas, fracciones,

gráficas y comprensión de operaciones) y un 33% (ítems 29, 30, 36 y 37) corresponde al contenido de **resolución de problemas**, que van del 29 al 40.

Tabla 10. Ítems no alineados

Ítems	Contenido intencional/propuesto (contenido de área) →Objetivos educativos/taxonomía	Contenido anticipado (procesos cognitivos) →Expertos en contenido/profesores	Resultado
Ítem 6 “B”	Conocimiento/recuerdo	Comprensión (síntesis) Utilización	No alineado
Ítem 8 “B”	Conocimiento/recuerdo	Comprensión (síntesis)	No alineado
Ítem 9 “B”	Conocimiento/recuerdo	Comprensión (síntesis) Utilización	No alineado
Ítem 18 “A y B”	Conocimiento/recuerdo	Comprensión (representación)	No alineado
Ítem 20 “B”	Comprensión	Conocimiento/recuerdo (nombrar)	No alineado
Ítem 25 “B”	Análisis	Comprensión (representación) Comprensión (síntesis)	No alineado

En la Tabla 10 se aprecia que seis ítems del total de los cincuenta ítems analizados, representan un 12% de “no alineación” de la prueba. De ellos, un 25% (ítems 6, 8 y 9) corresponden al contenido de **operaciones**, (ítems del 1 al 12) que incluyen suma, resta, multiplicación y división; y un 19% (ítems 18, 20 y 25) se refieren al contenido de **conceptos aritméticos**, (ítems del 13 al 28), es decir numeración, geometría, medidas, fracciones, gráficas y comprensión de operaciones. No existen ítems “no alineados” en cuanto al contenido de **resolución de problemas**.

Tabla 11. Ítems sin evidencia de alineación

Ítems	Contenido intencional/propuesto (contenido de área) →Objetivos educativos/taxonomía	Contenido anticipado (procesos cognitivos) →Expertos en contenido/profesores	Resultado
Ítem 21 “A y B”	Comprensión		No existe evidencia
Ítem 24 “A y B”	Conocimiento/recuerdo		No existe evidencia
Ítem 27 “A y B”	Análisis		No existe evidencia
Ítem 28 “A y B”	Comprensión		No existe evidencia

En la Tabla 11 se muestran los cuatro ítems del total de los cincuenta ítems analizados, en donde no se encontró evidencia de contenido y representan un 8% de ítems que quedaron fuera del análisis, ya que no se contaba con el registro de respuestas proporcionadas por los niños.

VI. DISCUSIÓN

El objetivo de la presente investigación se centró en evaluar la validez predictiva, de constructo y del nivel cognitivo de los ítems de una prueba de matemáticas de tercer grado que utiliza el Centro de Investigaciones Educativas –CIE- de la Universidad del Valle de Guatemala. En los resultados sobresale que, en cuanto a la validez predictiva, se encontró un coeficiente R^2 de 0.417 ($r=0.646$, $p<0.01$) entre las puntuaciones de la prueba de matemáticas de tercer grado y las puntuaciones obtenidas por los mismos niños en la prueba de matemáticas de sexto grado, lo cual quiere decir que los resultados de la prueba de matemática de sexto grado dependen en un 41.7% de los resultados de la prueba de matemática de tercer grado. Esto implica que existe consistencia entre las puntuaciones obtenidas por los mismos niños en ambas pruebas, lo cual es favorable en el sentido de poder utilizar la prueba de tercer grado como un buen predictor sobre cómo se encontrará su nivel de matemáticas cuando los niños lleguen a sexto grado. De ello se deriva que dichos resultados pueden ser tomados en cuenta para tomar decisiones respecto a cómo mejorar la calidad del aprendizaje de la matemática, puesto que si un niño obtiene un resultado “bajo” en tercer grado, según la evidencia encontrada, se espera que su resultado sea muy parecido en sexto grado. Por lo cual es importante tomar acciones preventivas, de manera que los aprendices logren un mejor aprovechamiento en torno al conocimiento y dominio de la matemática en el transcurso de los años. Y esto iría de acuerdo con la definición de validez proporcionada por Messick en los años setenta, sobre la incorporación al análisis de la validez, de las consecuencias sociales del uso de las pruebas como una fuente más de evidencia dentro del proceso de validación (Padilla *et al.*, 2006). Y dentro del contexto educativo, según Nunnally (1987), la validez predictiva tiene mayor importancia en las pruebas escolares que sirven para determinar los métodos de estudio idóneos.

Asimismo, estos resultados concuerdan con los hallazgos del estudio de Ramírez et al. (2013) que encontró un fuerte valor predictor del instrumento utilizado después de tres años de aplicado el mismo. Así como el estudio de Toro et al. (2014), quienes encontraron una adecuada validez de criterio ($r=.656$, $p<.01$). Es decir que, la evidencia hallada en el presente trabajo de investigación apoya el uso del criterio (resultados de la prueba de

matemáticas de sexto grado) como fuente de validez predictiva de la prueba de matemáticas de tercer grado y con ello se respalda la definición sobre que el criterio puede ser definido como el modelo contra el cual se compara y evalúa la puntuación de una prueba (Cohen y Swerdlik, 2006).

En cuanto a la evidencia de validez de constructo, en la presente investigación se encontró, mediante el análisis factorial confirmatorio con el método de extracción de componentes principales sobre cuatro factores (conocimiento, comprensión, análisis, utilización), de acuerdo a la taxonomía de Marzano, que el porcentaje de la varianza total explicada por los resultados de la prueba de matemática de tercer grado es de 45.6%, de esta forma: 15.98% en el factor 1, 11.62% en el factor 2, 9.92% en el factor 3 y 8.08% en el factor 4. El análisis factorial es utilizado para identificar factores o variables específicas que de manera particular son atributos, características o dimensiones en las que es posible diferir entre los sujetos de estudio (Cohen y Swerdlik, 2006).

Y según estos mismos autores, la denominación de factores que surgen de un análisis factorial tiene más que ver con el conocimiento, el juicio y la abstracción verbal y no hay reglas específicas para designar factores, sino solamente se ejerce un juicio concerniente a qué nombre del factor común comunica mejor su significado. Y tal como menciona Morales (2013) es necesario además hacer un análisis conceptual y ver qué tienen en común las formulaciones de los ítems que definen el mismo factor, sobre todo las de los ítems que tienen los pesos mayores, que son los que mejor lo definen. Y después de este análisis hay que dar un nombre a estos factores.

De tal forma que se podría mencionar que los ítems que se agrupan en el factor 1, tienen en común el *conocimiento* básico sobre operaciones de suma, resta, multiplicación, geometría, fracciones y medidas, ya que requieren no solo cierto nivel de reconocimiento, sino además la producción de información almacenada (Marzano y Kendall, 2008). Asimismo se podría deducir un nivel de *comprensión* de geometría, fracciones y numeración en los ítems que se agrupan en el factor 2, puesto que envuelven extraer el conocimiento hacia características clave organizadas en una forma generalizada como

información y eventos generales (Marzano y Kendall, 2008). En los ítems que se agrupan en el factor 3, adicional al conocimiento básico de operaciones como resta y multiplicación, se podría decir que involucran un nivel de *utilización* de este conocimiento, proceso que es utilizado por el individuo cuando desea completar una tarea específica como seleccionar entre dos o más alternativas, identificar los criterios que se usan al juzgar el valor de cada alternativa y por último, seleccionar entre las alternativas basadas en la aplicación del criterio (Marzano y Kendall, 2008). Y los ítems que se agrupan en el factor 4, se podría señalar que incluyen el *análisis* tanto de figuras geométricas como en la resolución de problemas, lo cual implica generar nuevas aplicaciones de una generalización o principio conocidos (Marzano y Kendall, 2008).

Respecto a lo anterior, esta es la parte medular del análisis factorial, ya que lograr visualizar el constructo que subyace detrás de los ítems que se tienden a agrupar en un mismo factor, es una tarea hasta cierto punto difícil para el investigador o quien elabora el instrumento de medición, puesto que implica tener conocimiento teórico sobre lo sustancial que mide la prueba, que en el caso de la presente investigación se facilita por tener al “alcance de la mano” los ítems que conforman tanto la forma A como la forma B de la misma y al ir analizando cada uno de ellos, fue posible “nombrar” los factores basados en la clasificación de la taxonomía de Marzano (ver tabla 7 en Resultados).

Por otra parte, de acuerdo a Morales (2013), lo que se busca mediante el análisis factorial es que entre todos los factores se explique la mayor proporción de varianza posible y aunque un menor número de factores o componentes expliquen menos varianza, la estructura es más clara e interpretable.

De esta manera los resultados encontrados en la presente investigación referente a los cuatro factores analizados (conocimiento, comprensión, análisis y utilización) que explican el 45.6% de la varianza, son similares a los hallazgos de Castrillón y Borrero (2005), en donde se determinó como resultado seis factores en el análisis factorial que explicaron el 51.74% de la varianza. Asimismo la investigación de González-Forteza et al. (2011), en el que el análisis factorial con extracción de componentes principales y rotación

oblicua arrojó una estructura de cuatro factores que explicaron 54.1% de la varianza. De igual forma concuerda con lo encontrado en el 2013 por Cervantes et al., en donde el análisis factorial exploratorio derivó inicialmente cuatro factores y se optó por replicar el análisis forzando la extracción a tres factores que explicaron 44.1% de la varianza; también guarda relación con el estudio de Romo et al. en 2014, en el que después del análisis factorial de componentes principales, los ítems se agruparon en cuatro factores que explicaron un total de 43.16% de la varianza. Y finalmente, tiene semejanza con el estudio en 2014, de Omar et al. en el que se estudió la validez factorial exploratoria y confirmatoria y los resultados mostraron la existencia de cuatro factores con satisfactorios niveles de validez que explicaron el 57% de la varianza total.

Con base en lo anterior es relevante mencionar que el análisis factorial es uno de los métodos que permiten llegar a establecer la validez de constructo de las puntuaciones de una prueba, pero no es el único método para lograrlo, ya que se pueden asimismo considerar otros, como es el método de validez convergente y discriminante, en donde se espera que los resultados de una prueba se relacionen con otra variable que se asume mide lo mismo, así como que difiera o no correlacione con variables con las que ha de diferenciarse. De igual forma, establecer diferencias entre grupos ayuda también a apoyar este tipo de validez. Pero en la presente investigación se decidió llevarlo a cabo por medio del análisis factorial, por ser un método ampliamente utilizado en estudios principalmente enfocados en el campo de la elaboración de instrumentos de medición.

En lo referente a la evidencia de validez del nivel cognitivo de los ítems de la prueba de matemáticas de tercer grado, en esta investigación se encontró un 60% de “alineación” total de los ítems analizados, entre lo planteado por los creadores de la misma según la tabla de especificaciones (ver Apéndices), los expertos en contenido (estudiantes del curso Investigación Cualitativa de las diferentes maestrías en educación de la Universidad del Valle de Guatemala) y las respuestas orales de los niños, que fueron grabadas y posteriormente transcritas para el respectivo análisis de contenido.

El método de validez cognitiva de ítems es útil para determinar la concordancia entre el juicio de expertos en contenido al analizar las respuestas proporcionadas por los niños mientras responden cada ítem de la prueba, utilizando para ello los protocolos de “pensamiento en voz alta” y comparar sus valoraciones con lo establecido por quienes elaboraron el instrumento, en este caso basado en el nivel de procesamiento cognitivo de acuerdo a la taxonomía de Marzano.

Al utilizar los protocolos de “pensamiento en voz alta” se obtiene evidencia de la forma como piensan y razonan los niños cuando resuelven cada ítem de la prueba y ello permite analizar el nivel de procesamiento cognitivo que está ejecutando el niño en ese momento y luego clasificarlo en el nivel correspondiente de la taxonomía. Esto con el fin de demostrar que la forma como fue concebido el instrumento de medición, en efecto coincide con el análisis del nivel cognitivo de los ítems que conforman el mismo.

Estos resultados muestran relación con el estudio de Barcelata *et al.* (2009) en el que se examinó la validez de contenido a través de la participación de jueces expertos y los análisis mostraron coeficientes de concordancia altos. Asimismo con el estudio de Toro *et al.* (2014), en el que se obtuvo la validez de contenido por medio de tres jurados expertos. Y con el estudio de McCorry *et al.* (2013) sobre validez de contenido utilizando el método “pensar en voz alta”. Finalmente, concuerda también con la investigación realizada por Renberg *et al.* (2008) en la que exploraron la validez de contenido, en donde los sujetos llenaron el instrumento mientras reportaban cómo razonaban en voz alta; las respuestas fueron grabadas, transcritas literalmente y analizadas y la combinación de los protocolos de “pensamiento en voz alta” y el análisis cualitativo informaron sobre la validez del instrumento y fue valioso en el desarrollo del cuestionario.

VII. CONCLUSIONES

Con base en la evidencia sobre las fuentes de validez de la prueba de matemática de tercer grado, se encontró que los resultados apoyan la validez predictiva de las puntuaciones de la prueba, ya que de acuerdo al coeficiente de correlación obtenido entre los resultados de la prueba de tercer grado y los resultados de los mismos niños en la prueba de sexto grado, se observa que dicho coeficiente es moderadamente alto, con un valor de 0.646 ($p < 0.01$) y un coeficiente de determinación (R^2) de 0.417, es decir las puntuaciones de la prueba de matemática de tercer grado son responsables en un 41.7% del resultado obtenido por los niños en la prueba de matemática de sexto grado. Dicha evidencia apoya lo planteado en la hipótesis alterna y por consiguiente se rechaza la hipótesis nula de investigación, ya que sí existe correlación estadísticamente significativa a nivel 0.05 entre los resultados de la prueba de matemáticas de tercer grado y los resultados obtenidos por los mismos niños en la prueba de matemáticas de sexto grado.

En lo referente a la validez de constructo, se realizó un análisis factorial confirmatorio, con el fin de contrastar los ítems planteados de acuerdo a la taxonomía de Marzano en la tabla de especificaciones de la prueba elaborada por el CIE (ver apéndices), con los resultados del análisis factorial. Al extraer cuatro factores, el porcentaje de la varianza explicada es de 45.6%.

Al analizar el primer factor se agrupan los ítems 1, 3, 4, 9, 14, 19, 23, 24, 26, 30, 33 y 36 (así como los ítems 8, 11 y 12 que comparten cargas factoriales mayores en el segundo factor). Este primer factor explica un 15.98% de la varianza total. En estos ítems se puede notar que algo en común es el *conocimiento* básico sobre operaciones de suma, resta, multiplicación, geometría, fracciones y medidas, lo cual implica transferir lo que se sabe a un estado actual de atención, tanto en forma de reconocimiento (nombrar), como de evocación o recuerdo (ejecutar), según lo mencionan Marzano y Kendall (2008). Según la tabla de especificaciones los ítems del nivel de conocimiento son: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 17, 18 y 24, de los cuales finalmente se recomienda re-clasificar los ítems 6, 8 y 18 a distintos niveles cognitivos (ver Tabla 12).

En un segundo factor se agrupan los ítems 2, 8, 10, 11, 12, 15, 18, 20, 27, 29 y 32 (así como los ítems 3, 16 y 31 que comparte cargas factoriales mayores en otros factores). Este segundo factor explica un 11.62% de la varianza total. Estos ítems conllevan el conocimiento básico de operaciones no solo de suma y multiplicación, sino además de división, así como un nivel de *comprensión* de geometría, fracciones y numeración, en donde este proceso es el responsable de identificar atributos críticos y definidos del conocimiento, de acuerdo a Marzano y Kendall (2008). Según la tabla de especificaciones los ítems del nivel de comprensión son: 13, 14, 19, 20, 21, 23 y 28, de los cuales se recomienda re-clasificar los ítems 14 y 28 a distintos niveles cognitivos (ver Tabla 12).

En un tercer factor se agrupan los ítems 5, 6, 7, 21, 28, 34, 37 y 38 (así como los ítems 19, 30, 32 y 33 que comparten cargas factoriales mayores en otros factores). Este tercer factor explica un 9.92% de la varianza total. En estos ítems, adicional al conocimiento básico de operaciones como resta y multiplicación, se involucra un nivel de *utilización* de este conocimiento y este proceso es utilizado por el individuo cuando desea completar una tarea específica (Marzano y Kendall, 2008). Según la tabla de especificaciones los ítems del nivel de utilización son: 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 y 40, de los cuales se recomienda re-clasificar el ítem 29 a distinto nivel cognitivo (ver Tabla 12).

Finalmente, en un cuarto factor se agrupan los ítems 13, 16, 22, 31, 35 y 39 (así como los ítems 5, 15, 21, 26, 28, 29, 30, 32 y 37 que comparten cargas factoriales mayores en otros factores). Este cuarto factor explica un 8.08% de la varianza total. Estos ítems incluyen el *análisis* tanto de figuras geométricas como en la resolución de problemas, lo cual implica la generación de nueva información que el individuo previamente no poseía, según Marzano y Kendall (2008). Según la tabla de especificaciones los ítems del nivel de análisis son: 15, 16, 22, 25, 26 y 27, de los cuales no se recomienda re-clasificar ningún ítem (ver Tabla 12).

En cuanto a los ítems que no mostraron cargas factoriales mayores a 0.30 en ninguno de los cuatro factores, se encontraron los ítems 17, 25 y 40, ya que para que un

ítem se considere que pertenece a un factor, o dicho con más propiedad, que lo define o explica suficientemente, debe tener en él una correlación razonablemente alta (por lo menos 0.30, o más) y no tenerlas mayores en otro factor (Morales, 2013).

Respecto a la validez del nivel cognitivo de los ítems, si coinciden el **contenido intencional** (propuesto en la tabla de especificaciones) y el **contenido anticipado** (basado en los juicios de expertos sobre los protocolos de “pensamiento en voz alta” de los niños) en su totalidad en cuanto al proceso cognitivo principal que refleja el ítem, entonces se dice que el ítem está “*alineado*”. De tal forma que, treinta ítems del total de los cincuenta ítems analizados están “*alineados*” en su totalidad, es decir representan un 60% de “*alineación*” de la prueba. De ellos seis ítems corresponden al nivel cognitivo de *conocimiento*, cuatro ítems al nivel de *comprensión*, seis ítems al nivel de *análisis* y catorce ítems al nivel de *utilización*.

Asimismo, si difieren el **contenido intencional** y el **contenido anticipado** en alguno de los procesos cognitivos que refleja el ítem, se dice que el ítem está “*parcialmente alineado*”. Se observó que diez ítems del total de los cincuenta ítems analizados se encuentran “*parcialmente alineados*” y representan un 20% de “*alineación parcial*” de la prueba. De ellos, cuatro ítems corresponden al nivel cognitivo de *conocimiento*, un ítem al nivel de *comprensión*, un ítem al nivel de *análisis* y cuatro ítems al nivel de *utilización*.

Por otra parte, si **ambos contenidos** difieren en su totalidad, se dice que el ítem está “*no alineado*”. De esta manera, seis ítems del total de los cincuenta ítems analizados se encuentran “*no alineados*” y representan un 12% de “*no alineación*” de la prueba. De ellos cuatro ítems corresponden al nivel cognitivo de *conocimiento*, un ítem al nivel de *comprensión* y un ítem al nivel de *análisis* y ningún ítem “*no alineado*” corresponde al nivel de *utilización*.

En cuatro ítems del total de los cincuenta ítems analizados de la prueba, no se encontró evidencia de contenido y representan un 8% de ítems que quedaron fuera del análisis, ya que no se contaba con el registro de respuestas proporcionadas por los niños.

Por lo anterior se puede notar que existe mayor coincidencia o “*alineación*” en el nivel cognitivo de *utilización*, entre lo planteado por los creadores de la prueba de acuerdo a la tabla de especificaciones, los expertos en contenido y las respuestas orales de los niños. Esto puede explicarse ya que en la prueba se muestra la instrucción: “*resuelve los siguientes problemas*” y uno de los subprocesos del nivel cognitivo “*utilización*” es precisamente la “*resolución de problemas*”. Además las respuestas que los niños proporcionaron, reflejan que están intentando lograr una meta, e identificando los obstáculos hacia la misma, así como las alternativas para alcanzar dicha meta, evaluando las alternativas y seleccionando y ejecutando las alternativas, tal como lo indican Marzano y Kendall (2008) para este nivel de procesamiento cognitivo.

De acuerdo al contenido intencional o propuesto (contenido de área), la mayor parte de ítems se encuentra concebido por los creadores de la prueba como referido al nivel de cognición de *conocimiento/recuerdo*, según la taxonomía de Marzano, y al momento de contrastarlo con el contenido anticipado (juicio de expertos) basado en las respuestas de los niños, algunos de los expertos los clasificaron en distinto nivel de procesamiento cognitivo. Por ejemplo, los ítems 6 y 8, se consideraron que reflejan un nivel de *utilización* y *comprensión*, respectivamente, que se refieren a cuando el individuo desea completar una tarea específica y extraer el conocimiento hacia características clave organizadas en una forma generalizada (información y eventos generales), según Marzano y Kendall (2008). Y similar resultado se presentó con relación a los contenidos de operaciones aritméticas y resolución de problemas, pero cabe destacar que en este último contenido, la “*alineación parcial*” fue la más baja, es decir hubo mayor coincidencia entre los contenidos propuestos y anticipados.

El análisis anterior permite determinar que en su mayoría (60%) los ítems de la prueba de matemáticas de tercer grado coinciden en la clasificación del nivel cognitivo de acuerdo a la taxonomía de Marzano, que refleja cada uno de ellos, por lo cual se puede asumir que los resultados muestran un nivel de validez cognitiva apropiado, tal como lo menciona Messick (1975) en cuanto a las razones teóricas de la consistencia observada de

las respuestas (acciones y procesos de pensamiento), que se puede obtener al preguntar a un grupo de estudiantes que “piensen en voz alta” mientras responden un conjunto de preguntas.

En la Tabla 12 se presentan todos los ítems de la prueba de matemática, de acuerdo a la clasificación dada en la tabla de especificaciones, así como la alineación del nivel cognitivo de cada uno de ellos, seguido de la clasificación a partir del análisis factorial y, basado en esta evidencia, se incluye la recomendación de reclasificar el nivel cognitivo de los ítems 6, 8, 14, 18, 28 y 29.

Tabla 12. Reclasificación de ítems

No. ítem	Tabla de especificaciones	Alineación cognitiva	Análisis Factorial	Recomendación
1	Conocimiento	Conocimiento	Conocimiento	Conocimiento
2	Conocimiento	Con/Uti	Comprensión	Conocimiento
3	Conocimiento	Conocimiento	Conocimiento	Conocimiento
4	Conocimiento	Con/Uti	Conocimiento	Conocimiento
5	Conocimiento	Conocimiento	Utilización	Conocimiento
6	Conocimiento	Com/Uti	Utilización	Utilización
7	Conocimiento	Conocimiento	Utilización	Conocimiento
8	Conocimiento	Comprensión	Comprensión	Comprensión
9	Conocimiento	Com/Uti	Conocimiento	Conocimiento
10	Conocimiento	Conocimiento	Comprensión	Conocimiento
11	Conocimiento	Con/Uti	Comprensión	Conocimiento
12	Conocimiento	Conocimiento	Comprensión	Conocimiento
13	Comprensión	Comprensión	Análisis	Comprensión
14	Comprensión	Con/Com	Conocimiento	Conocimiento
15	Análisis	Análisis	Comprensión	Análisis
16	Análisis	Análisis	Análisis	Análisis
17	Conocimiento	Con/Uti	(sin carga factorial)	Conocimiento
18	Conocimiento	Comprensión	Comprensión	Comprensión
19	Comprensión	Comprensión	Conocimiento	Comprensión
20	Comprensión	Conocimiento	Comprensión	Comprensión
21	Comprensión	Sin evidencia	Utilización	Comprensión
22	Análisis	Análisis	Análisis	Análisis
23	Comprensión	Comprensión	Conocimiento	Comprensión
24	Conocimiento	Sin evidencia	Conocimiento	Conocimiento
25	Análisis	Comprensión	(sin carga factorial)	Análisis
26	Análisis	Análisis	Con/Ana	Análisis
27	Análisis	Sin evidencia	Comprensión	Análisis
28	Comprensión	Sin evidencia	Utilización	Utilización
29	Utilización	Com/Uti	Comprensión	Comprensión
30	Utilización	Uti/Ana	Con/Ana/Uti	Utilización
31	Utilización	Utilización	Análisis	Utilización
32	Utilización	Utilización	Com/Uti	Utilización
33	Utilización	Utilización	Con/Uti	Utilización
34	Utilización	Utilización	Utilización	Utilización
35	Utilización	Utilización	Análisis	Utilización
36	Utilización	Uti/Ana	Conocimiento	Utilización
37	Utilización	Uti/Ana	Utilización	Utilización
38	Utilización	Utilización	Utilización	Utilización
39	Utilización	Utilización	Análisis	Utilización
40	Utilización	Utilización	(sin carga factorial)	Utilización

VIII. RECOMENDACIONES

Basado en los hallazgos de la presente investigación se recomienda:

- Reclasificar el nivel cognitivo del ítem 6 al nivel de “utilización” con base en la evidencia del análisis factorial, así como en el resultado de la alineación parcial de dicho ítem.
- Reclasificar el nivel cognitivo del ítem 8 al nivel de “comprensión”, basado en la evidencia del análisis factorial y el juicio de expertos fundamentado en las respuestas orales de los niños.
- Reclasificar el nivel cognitivo del ítem 14 al nivel de “conocimiento” con base en la evidencia del análisis factorial, así como en el resultado de la alineación parcial de dicho ítem.
- Reclasificar el nivel cognitivo del ítem 18 al nivel de “comprensión”, basado en la evidencia del análisis factorial y el juicio de expertos fundamentado en las respuestas orales de los niños.
- Reclasificar el nivel cognitivo del ítem 28 al nivel de “utilización” con base en la evidencia del análisis factorial y el juicio de la responsable de la presente investigación.
- Reclasificar el nivel cognitivo del ítem 29 al nivel de “comprensión” con base en la evidencia del análisis factorial, así como en el resultado de la alineación parcial de dicho ítem.

IX. BIBLIOGRAFÍA

- Aiken, L. (2003). *Test psicológicos y evaluación* (11va. ed.). México: Prentice Hall.
- Alvarado, J. y Santisteban, C. (2006). *La validez en la medición psicológica*. Madrid: UNED ediciones.
- Aragón, L. (2011). *Evaluación psicológica: historia, fundamentos teórico-conceptuales y psicometría*. México D.F., México: Manual Moderno.
- Armas, S. (2002). *Validación de los instrumentos de evaluación docente en la Facultad de ciencias Químicas y Farmacia de la Universidad de San Carlos de Guatemala*. Tesis de maestría, Universidad del Valle de Guatemala.
- Armengol, L. (2007). Los protocolos de pensamiento en voz alta como instrumento para analizar el proceso de escritura. *RESLA*, 20, 27-35.
- Barcelata, B., Gómez, E. y Durán, C. (2009). Construcción, confiabilidad, validez de contenido y discriminante del Inventario Autodescriptivo del Adolescente. *Acta Colombiana de Psicología*, 9 (2), 5-17.
- Casas, L. y Luengo, R. (2004). Representación del conocimiento y aprendizaje. Teoría de los conceptos nucleares. *Revista española de pedagogía*, 227, 59-84.
- Castrillón, D. y Borrero, P. (2005). Validez estructural y confiabilidad del inventario de ansiedad estado – rasgo (STAIC) para la valoración de la ansiedad en niños. *Suma psicológica*, 12 (1), 47-60
- Cervantes, B., Álvarez, G., Gómez, G. y Mancilla, J. (2013). Confiabilidad y validez de constructo de la Escala de Modificación Corporal en varones adolescentes. *Salud Mental*, 36 (6), 487-492.
- Cohen, R. y Swerdlik, M. (2006). *Pruebas y evaluación psicológicas* (6a. ed.). México D.F., México: McGraw-Hill Interamericana.
- Cook, D. y Beckman, T. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine*, 119 (2), 166.e7 – 166.e16.
- Cruz, Ch., Díaz-Loving, R., Oropeza, R., Díaz, P., Padilla, J. y González, I. (2013). Temor a la evaluación social negativa: validez de constructo y criterio del instrumento de medición. *Universitas Psychologica*, 12 (2), 531-545.

- De la Fuente, S. (2011). *Análisis factorial*. Universidad Autónoma de Madrid, España.
- Elosúa, P. (2003). Sobre la validez de los tests. *Psicothema*, 15 (2), 315-321.
- Escartín, J., Rodríguez-Muñoz, A., Sora, B. y Rodríguez-Carballeira, A. (2012). Adaptación y validación de la versión española de la Escala de Conductas Negativas en el Trabajo realizadas por los acosadores. *Revista de Psicología del Trabajo y de las Organizaciones*, 28 (3), 157-170.
- Fortín, A. (1998). *Estudio de la validez concurrente de los instrumentos elaborados por las maestras de una institución privada mixta de la ciudad de Guatemala para evaluar el aprestamiento a la lectura*. Tesis de maestría, Universidad del Valle de Guatemala.
- Gálvez-Sobral, J. (2008). *Desarrollo y validación de una escala para medir bullying en alumnos de sexto primaria del municipio de Guatemala*. Tesis de licenciatura, Universidad Rafael Landívar, Guatemala.
- Gómez, J. e Hidalgo, M. (s.f.). *La validez en los tests, escalas y cuestionarios*. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro de Estudios de Opinión.
- Gonzáles, C., López, L. y Montenegro, H. (2006). Análisis de confiabilidad y de validez del instrumento Course experience questionnaire (CEQ). *Educ. Educ.*, 15 (1), 63-78.
- González-Forteza, C., Solís, C., Jiménez, A., Hernández, I., González, A., Juárez, F., Medina, M. y Fernández, H. (2011). Confiabilidad y validez de la escala de depresión CES-D en un censo de estudiantes de nivel medio superior y superior, en la ciudad de México. *Salud Mental*, 34 (1), 53-59.
- Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la investigación* (5ª. ed.). México D.F., México: McGraw-Hill.
- Kerlinger, F. y Lee, H. (2002). *Investigación del comportamiento*. México: McGraw-Hill.
- Márquez-Sandoval, Y., Salazar-Ruiz, E., Macedo Ojeada, C., Altamirano-Martínez, M., Bernal-Orozco, M., Salas-Salvado, J. y Vizmanos-Lamotte, B. (2014). Diseño y validación de un cuestionario para evaluar el comportamiento alimentario en estudiantes mexicanos del área de la salud. *Nutrición hospitalaria*, 30 (1), 156-164.

- Marzano, R. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin.
- Marzano, R. y Kendall, J. (2008). *Designing and assessing educational objectives: applying the new taxonomy*. Thousand Oaks: Corwin Press.
- McCorry, N., Scullion, L., McMurray, C., Houghton, R. y Dempster, M. (2013). Content validity of the illness perceptions questionnaire – revised among people with type 2 diabetes: A think-aloud study. *Psychology and Health, 28* (6), 675-685.
- Messick, S. (1975). Meaning and values in measurement and evaluation. *American Psychologist, 30*, 955-966.
- Morales, P. (2011). *Guía para construir cuestionarios y escalas de actitudes*. Guatemala: Cindeg, Facultad de Humanidades, Universidad Rafael Landívar.
- Morales, P. (2013). *El análisis factorial en la construcción e interpretación de tests, escalas y cuestionarios*. Recuperado de: <http://www.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf>
- Newton, P. (2013). *Does it matter what 'validity' means?* Leading education and social research. Institute of Education, University of London.
- Nunnally, J. (1987). *Teoría psicométrica*. México, D.F: Trillas.
- Omar, A., Salessi, S. y Urteaga, F. (2014). Diseño y validación de la escala Capps para medir capital psicológico. *Liberabit, 20* (2), 315-323.
- Padilla, J., Gómez, J., Hidalgo, M. y Muñiz, J. (2006). La evaluación de las consecuencias del uso de los tests en la teoría de la validez. *Psicothema, 18* (2), 307-312.
- Ramírez, Y., Díaz, M., Ramos, F. y Manga, D. (2013). Validez y confiabilidad de la Batería Luria Inicial para identificar alteraciones neuropsicológicas en niños cubanos. *Revista cubana de neurología y neurociencia, 3* (1), 18-25.
- Renberg, T., Lindblad, A. y Tully, M. (2008). Testing the validity of a translated pharmaceutical therapy-related quality of life instrument, using qualitative 'think aloud' methodology. *Journal of clinical pharmacy and therapeutics, 33*, 279-287.
- Ritchey, F. (2008). *Estadística para las ciencias sociales* (2ª. ed.). México, D.F.: McGraw Hill.

- Romo, T., Enríquez, C., Hernández, M., Ruiz, M., Castillo, R., Ehrenzweig, Y., Marván, M. y Larralde, C. (2014). Validación en México del inventario de ajuste de Weinberger (WAI). *Salud Mental*, 37, 247-253.
- Soto, H. (2007). *Modelo predictivo del rendimiento académico en estudiantes repitentes, de primer año Facultad de Ciencias Médicas Universidad de San Carlos de Guatemala*. Tesis de maestría, Universidad del Valle de Guatemala.
- Thelk, A., Hoole, E. y Lottridge, S. (s.f.). *What are you thinking? Postsecondary student think-alouds of scientific and quantitative reasoning items*. Cognitive validity of test items and scores. James Madison University.
- Tornimbeni, S., Pérez, E. y Olaz, F. (2008). *Introducción a la psicometría*. Buenos Aires: Paidós.
- Toro, R., Arias, A. y Quant, D.M. (2014). Diseño y validación del Cuestionario de Autoesquemas Desadaptativos en Fobia Social CAD-FS. *Revista Colombiana de Psicología*, 23 (1), 135-148.

X. APÉNDICES

Tabla de especificaciones de la prueba de matemáticas de tercer grado

No.	Contenidos	No. del ítem	Niveles de Marzano			No. total	Por ciento	
			Conocimiento	Comprensión	Análisis			Utilización
1	Operaciones		37.50%	17.50%	15.00%	30.00%	12	30.0%
	a. Suma	1,2,3	1,2,3				3	7.5%
	b. Resta	4,5,6	4,5,6				3	7.5%
	c. Multiplicación	7,8,9	7,8,9				3	7.5%
	d. División	10,11,12	10,11,12				3	7.5%
2	Conceptos Aritméticos						14	35.0%
	a. Numeración	20,21		20,21			2	5.0%
	b. Geometría	13,14,15,16		13,14	15,16		4	10.0%
	c. Medidas	17, 23,24	17,24	23			3	7.5%
	d. Fracciones	18,19	18	19			2	5.0%
	e. Gráficas	22			22		1	2.5%
	f. Comprensión operaciones	25,26,27,28		28	25,26,27		4	10.0%
3	Resolución de problemas						12	30.0%
	a. Suma	29,30				29,30	2	5.0%
	b. Resta	31				31	1	2.5%
	c. Multiplicación	35				35	1	2.5%
	d. División	32,37				32,37	2	5.0%
	e. Varias operaciones	33,34,36,38				33,34,36,38	4	10.0%
	f. Lógica	39,40				39,40	2	5.0%
TOTALES							40	100

Fuente: Centro de Investigaciones Educativas, Universidad del Valle de Guatemala.

Contenido actual (estrategias de respuestas de los examinados)

Protocolos de pensamiento en voz alta de los niños

<p>Ítem 1 Forma A</p>	<p>“876 más 413... 6 más 3, escribo?... 9. 7 más 1 igual 8. 8 más 4 igual 12.... Esta ve, la d”.</p> <p>“8 más 4 igual 10... 12, 12... 8 más 4, 7 más 1, 6 más 3.... 6 más 3 (utiliza sus dedos para contar)... 9... 7 más 1 igual 8... 8 más 4 igual 12... 1,289”.</p> <p>“8 más 4... 7 más una... 6 más 3... Contar... Eee 8 más 4... 7 más una igual 8... 6, 6 más 3 sería... 9... 289... No porque el uno lo tengo que poner aquí... 1289... la d... Subrayarla”.</p> <p>“Entonces 876 más 413, 6 más 3 es 9, 7 más 1 es 8 y 8 más 4 es 12. Y es 1289”.</p>
<p>Ítem 2 Forma B</p>	<p>“Dos más dos... cuatro... cinco más cuatro, nueve... seis más cinco... once”.</p> <p>“Cinco más... cinco más cuatro, me da cosa... cinco más cuatro más cero, son nueve. Tres más dos más seis son... tres más dos son... cinco. Cinco más seis son once”.</p> <p>“Uno más uno... dos... cuatro... nueve... sumé cinco con el cuatro... nueve... igual que aquí... da cinco... sumé tres con el dos... doce, me da doce, mmm... once”.</p> <p>“Mmm, aquí voy a sumar la cantidad de trescientos cincuenta y uno más doscientos cuarenta y uno más seiscientos dos... Y la respuesta es mil ciento noventa y cuatro... Porque sumé las cantidades... Y me salió el resultado correcto”.</p> <p>“Uno más uno más dos es cuatro, aquí pongo el número también?... luego cinco más cuatro es nueve, tres más dos es cinco más seis es onceaquí está, la respuesta es la b”.</p> <p>“Cinco más cuatro nueve, es nueve, como aquí hay un cero no lo puedo sumar entonces cinco más cuatro es nueve y pongo el nueve acá y el tres más dos cinco más seis es once, entonces serían mil ciento noventa y cuatro y miiii mi respuesta sería esta, la b”.</p> <p>“Ahí emm mmm yo vine y empecé a sumar dos más uno más uno me dio cuatro... entonces ya que puse el cuatro aquí como aquí hay un cero vine y sumé cinco más cuatro y me dio nueve luego vine y sumé seis más tres me dio nueve más dos me dio me dio once y el resultado es ciento, mil ciento noventa y cuatro”.</p>
<p>Ítem 3 Forma A</p>	<p>“Sumaría dos más uno, nueve más dos, siete más tres y cuatro más seis”.</p> <p>“Lo haría desde la derecha hacia la izquierda... seis más cuatro, diez... voy a sumar tres más siete, esto da once... aquí era trece, no, aquí era doce pero le sumé el de arriba y lo puse al de al lado”.</p>

	<p>“Seis más cuatro... me da diez, entonces pongo el cero y me llevo uno... entonces en lugar de ser siete más tres, serían siete más cuatro; entonces siete más cuatro me da once... me lo llevo al dos, entonces en lugar de ser nueve por dos, nueve más dos, sería nueve más tres, doce me daría, entonces dejo el dos y me llevo el uno, y en lugar de ser dos más uno, es dos más dos y cuatro”.</p> <p>“Aquí empiezo a sumar... el seis más el cuatro... diez y llevo una... aquí sumo uno más tres que me da cuatro, y cuatro más siete me da once... llevo una y aquí pongo el resto... sumo uno más dos que me da tres y tres más nueve que me da once... llevo una y aquí pongo el resto... sumo uno más dos que me da tres y tres más nueve que me da once, aquí llevo uno y aquí el resto. Y aquí sumo uno más uno que me da dos y dos más dos me da cuatro”.</p>
<p>Ítem 4 Forma B</p>	<p>“La respuesta de la segunda es quinientos cuarenta y tres... lo hice verticalmente y me salió este resultado...”</p> <p>“Siete menos cuatro es tres, emmm seis menos dos es cuatro y ocho menos tres es cinco... es quinientos cuarenta y tres y la respuesta es la C”.</p> <p>“Siete menos cuatro esss, esss, no seee... seis menos dos es cuatro y ocho menos tres es cinco, mi respuesta sería quinientos cuarenta y tres, entonces sería la respuesta C”.</p> <p>“Ahí resté a siete le quité cuatro y me sobraron tres... Y acá puse a seis le quito dos me sobraron cuatro y a ocho le quito tres me sobraron cinco y la respuesta es quinientos cuarentitrés”.</p>
<p>Ítem 5 Formas A y B</p>	<p>“Restaría cinco menos seis... me daría uno... restaría nueve menos seis... seis menos nueve... me daría tres... pongo el ocho porque no hay nada que resta ahí”.</p> <p>“Restar... por ejemplo acá tengo que restar ochocientos noventa y seis menos sesenta y cinco... entonces tengo que buscar la respuesta y subrayarla... entonces si a seis le quito cinco me quedaría uno... si a nueve le quito seis como no se puede, el ocho le presta uno al nueve, entonces sería diecinueve más seis, entonces sería catorce... pongo el cuatro y me llevo el uno y este queda como siete y bajo el ocho... entonces sería ochocientos...”</p> <p>“Resto seis menos cinco que me da uno... resto nueve menos seis que me da tres y aquí solo bajo el ocho... porque se queda solito”.</p> <p>“6 menos 5 igual 1. 9 menos 6 igual 3... me quede pensando para restar... Aja 3. Y solo bajamos el 8... Porque ya no hay más números para poner, para restar. (balbucea) 831, está la a)”.</p> <p>“Tengo 6 y le quito 5 igual 1... Tengo 9 y le resto, 6..... 1, 2, 3, 4, 5, 6, 3....</p>

	<p>(utiliza los dedos)... Y bajo el 8... lo bajo... porque no tiene un número aquí... a) (pausa) 831”.</p> <p>“Eee 8 menos 0 igual 0... 9 menos 6... 3... 6 menos 5 igual una (uso de dedos)... (pausa) 0 y el 31”.</p> <p>“Entonces 6 menos 5 es 1, 9 menos 6 es 3 y 8 menos nada es 8. Es 831”.</p>
<p>Ítem 6 Forma B</p>	<p>“Ocho menos nueve... uno... el cuatro le presta... nueve menos nueve... mmm... dieciocho... quitarle nueve... nueve... Tres... siete menos cuatro... siete menos tres... no... el dos le presta al cuatro... uno... trece... mmm... lo mismo... trece... quitarle siete... seis... mmm... diez... cuatro... tres”.</p> <p>“Porque el cuatro le presta uno al ocho entonces se convierte en dieciocho, entonces le resto nueve menos dieciocho y que son nueve... como el cero no le puede prestar al cuatro entonces el dos le presta al cero se convierte en diez, pero como él le presta al cuatro se convierte en nueve yyy, no, entonces el tres se convierte en tres... trescientos sesenta y nueve”.</p> <p>“Ah mmm, primero tuve que restar ocho menos nueve pero como no se puede le presto uno al cuatro y ahora es de dieciocho menos nueve que... después como el cuatro se quedó en valor de tres emm, se le pide prestado uno al cero pero como el cero no tiene le va a pedir prestado al dos entoes el dos se lo da al cuatro entonces catorce menos siete y da siete después emm en el dos se quedó en uno le presta el último que tiene a la cero otra vez después diez menos seis son mm da cuatro y después solo baja el uno y queda cuatrocientos setenta y nueve”.</p> <p>“Ocho menos nueve no se puede , cuatro menos siete no se puede entonces en el dos menos uno le quito uno al dos se vuelve uno , el dos le presta uno al cero se vuelve nueve , es el cuatro con el cero le prestó uno al cuatro se queda en nueve, el cuatro se vuelve trece porque le presta uno al ocho y el ocho se vuelve dieciocho , luego dieciocho menos nueve es nueve trece menos siete es seis , nueve menos cuatro... nueveeee meenoos cuatro, esss tres y uno menos uno es cero así que allí no pongo nada así que la respuesta es trescientos sesenta y nueve es la d”.</p> <p>“Ocho menos nueve no se puede entonces reagrupó, el cuatro lo pongo después de éste y éste se queda en dieciocho, dieciocho menos nueve esss nueve, tres menos siete no se puede entonces voy a reagrupar otra vez pero como en el cero no se puede, lo reagrupó con el dos y queda en uno, el cero se queda en nueve y se puede poner para acá uno, entonces el tres se queda en trece, trece menos siete esss seis y nueve menos seis ess esss tres. Entonces pongo el tres, uno menos uno es cero</p>

	<p>entonces sería trescientos sesenta y nueve entonces sería la respuesta d”.</p> <p>“Pues aquí si me cuesta porque no me recuerdo muy bien entonces yo me recuerdo que le tiene que (prestar) ocho tiene menos y se le tiene que restar nueve entonces lo convertí a dieciocho porque le prestó un uno... Se le quita nueve y da nueve... Y aquí ya no me recuerdo entonces tengo que contar con cosas así... Es que yo ponía lapiceros para... entonces por eso pero ahí ya no me recuerdo muy bien por eso estoy pensando..... ahí está... me dio nueve otra vez porque como mmmm este y como me imaginé que tenía otra mano a la par entonces... restando nueve... me queda nueve...”</p>
<p>Ítem 7 Forma A</p>	<p>“Hice multiplicación porque dos por cuatro ocho... otra vez... y dos por siete catorce”.</p> <p>“Tengo que poner, tengo que hacer el doble de este número... me daría ocho... aquí también ocho... y aquí diez... aquí me daría catorce... aquí pondría que sería mil cuatrocientos ochenta y ocho”.</p> <p>“Dos por cuatro, me daría ocho... el otro igual, dos por cuatro, me da ocho... el otro dos por siete me daría catorce... entonces serían ciento... mil cuatrocientos ochenta y ocho”.</p> <p>“Multiplico cuatro por dos que me da ocho... aquí también me da ocho porque también hay un cuatro y es por dos... y aquí catorce”.</p>
<p>Ítem 8 Forma B</p>	<p>“Mmm, cuatro por... treinta y dos... pongo el dos y el... cuatro por siete... cuarenta y ocho... mmm... veintiocho... treinta y uno... sumar... tres”.</p> <p>“Cuatro por ocho son... treinta y dos. Y cuatro por siete son... veinte y ocho más tres son... treinta y uno”.</p> <p>“Cuatro por siete... trei... veintinueve, veintiocho... treinta y uno, le sumé estos tres”.</p>
<p>Ítem 9 Forma B</p>	<p>“Cinco por nueve... cuarenta y cinco... el cuatro... nueve por cuatro... treinta y seis... cuarenta... nueve por uno. Nueve... lo sumo... trece”.</p> <p>“Nueve por cinco son cuarenta y cinco. Nueve por cuatro son... treinta y seis. Treinta y seis más cuatro... seis más cuatro son diez. Treinta y seis, cuarenta. Cierto. Uno más... son trece”.</p> <p>“Nueve por cinco... cincuenta... treinta y cinco... De cinco en cinco... veinte, treinta, cuarenta, cincuenta, sesenta, setenta, ochenta... cinco, diez, quince, veinte, veinticinco, treinta, treintaicinco, cuarenta, cuarentaicinco... cuarentaicinco... voy... el cinco aquí... el cuatro... nueve por cuatro... el seis aquí... mmm... cuarenta...</p>

sumé estos cuatro... ¿ah?... quedan... treinta y seis... pero... aquí el cero... nueve por una es nueve, y es once, doce, ¡trece!”.

“Multiplicando nueve por cinco que serían cuarenta y cinco, entonces pongo el cuatro aquí , luego multiplico nueve por cuatro que serían treinta y seis más cuatro ya serían cuarenta, entonces pongo cero y pongo el cuatro aquí , entonces nueve por uno sería nueve entonces le sumo cuatro y serían trece... estee, sería, mil trescientos cinco”.

“Emm de primero nueve por cinco después nueve por cuatro y por último nueve por uno... mmm cuarenta y cinco... mm llevo cuatro... Mmmm la respuesta es mil trescientos cinco”.

“Cinco por nueve cuarenta y cinco, nueve por cuatro es nueve por cuatro es treinta y seis más cuatro cuarenta, nueve por nueve por una nueve más cuatro es trece entonces es mil trescientos cinco la respuesta es c... porque cinco como es cuarenta y cinco , el cinco tiene que estar aquí y qué hacemos con el cuatro? Lo subimos acá y lo sumamos con el siguiente número...”

“Vaa, nueve por cinco es cuarenta y cinco, pongo el cuatro y me queda el cinco, nueve por cuatro es es, noooo jijijiji... nueve por cinco... y después me dio cuarenta y cinco pero ahorita me estaba poniendo a pensar esss sería treinta y seis pero como aquí puse el cuatro tengo que sumar... treinta y seis más cuatro serían cuarenta, entonces aquí pongo cuatro... nueve por una sería nueve más cuatro serían trece, entonces serían mil trescientos cinco... y la respuesta sería la c”.

“Aquí voy a multiplicar cinco por nueve o nueve por cinco... Entonces multiplico cinco por nueve me da cuarenta y cinco pongo acá un cinco y subo el cuatro tonces acá tengo que multiplicar a nueve multiplicado por cuatro entonces nueve por cuatro por una cuatro, cuatro por dos; ocho, cuatro por tres doce, cuatro por cuatro dieciséis, cuatro por cinco; veinte, cuatro por seis; veinticuatro, cuatro por siete; veintiocho, cuatro por ocho treinta y dos, cuatro por nueve, treinta y seis, entonces ahora a multiplicar el uno lo bajo acá y entonces lo que me dio treinta y seis tengo que contar el de acá arriba lo sumo la multiplicación, la respuesta me dio treinta y seis porque cuatro por nueve da treinta y seis más cuatro: treinta y siete, treinta y ocho, treinta y nueve, cuarenta, aquí si subo el cuatro, acá pongo el cero, multiplico cuatro por una cuatro y mi respuesta me dio cuatrocientos cinco. Ah es que es que a mí me lo enseñaron así multiplico por ejemplo no importa si es nueve por cinco, cinco por nueve entonces la respuesta es la misma entonces pongo cuarenta y cinco pero aquí

	<p>solo ... el de abajo tengo que poner el cinco y arriba del otro número tengo que multiplicar tengo que poner el cuatro el número me da hasta acá cuarenta y cinco y entonces el nueve lo tengo que multiplicar por cuatro pero tengo que contar el de arriba entonces me dio cuarenta y el cuatro lo multiplico por uno me dio cuatro y me dio cuatrocientos cinco”.</p>
<p>Ítem 10 Forma A</p>	<p>“Tengo que buscar un número que se acerque al 18... 3 por 1 igual 3; 3 por 2 igual 6; 3 por 3 igual 9; 3 por 4 igual 12; 3 por 4 igual 12; 3 por 5 igual 15; 3 por 6 (pausa larga) igual 18... Me confundí, puse el 6 y no es el 6 es al 8.... Ahora resto y 0 y 0... La d”.</p> <p>“6... Porque 6 por 3 igual 18.... d”.</p> <p>“Tengo que poner un numero acá arriba para que me dé 18... Hmmm... El 5... Me da otro número... hmmm... 3 por una igual 3, 3 por 2 igual 6, 3 por 3 igual 9, 3 por 4 eeeee.... Allí es donde más me confundo porque... 3 por 4 igual 13... 3 por 5 igual 16... 3 por 6 igual 18... Subrayar el 6... la d”.</p> <p>“3 dividido 18, es 6. Porque 6 por 3 es 18... cabe 3 en 18”.</p>
<p>Ítem 11 Forma B</p>	<p>“Ahorita voy a multiplicar siete por seis, jijijiji, hayyyy, es que me confundoo, jijijiji...entonces serían siete por cinco son treinta y cinco por seis serían cuarenta y dos, entonces sería a seis, allí está... emmm, como vi que siete por seis es cuarenta y dos puse el seis arriba, puse cuarenta y dos abajo, eeem, resté y vi si no me quedaba residuo”.</p> <p>“Ahh tengo que buscar un número en la tabla del siete que me de cuarenta y dos y el número que me da cuarenta y dos es seis... Y el resultado es seis”.</p> <p>“La operación once que es siete dividido cuarenta y dos y es la respuesta a, cuarenta y dos le resto a cuarenta y dos que es cero y allí sí queda la operación”</p> <p>“Ahhhh, esta es división verdad... sí porque está la casita... sí división. Sieteete, sieteete no cabe en cuatro entonces sería, siete tiene que caber en cuarenta y dos entonces , jijijijiji, es que no sé... entonces sieteete, jijijijiji, ya me hice bolas... como aquí cabeee, a no perate, entonces aquí sería siete y cuánto caben en el cuarenta y dos y siete y cuarenta y dos cabe seis veces, aquí hay que poner seis, siete por seis es cuarenta y dos, aquí pongo cuarenta y dos y aquí tengo que restar, dos menos dos es cero y cuatro menos cuatro es cero entonces sería seis”.</p> <p>“Aquí si con estas creo que si no; es que a mí me cuesta mucho la división... siete dividido el siete da cuarenta y dos, a mí me lo enseñaron así, aquí tengo que restar, tonces pongo aquí em tengo que multiplicar siete por cuatro porque....siete.....</p>

	<p>siete.....por cuatro, aquí esta ve entonces aquí como a cuatro le quito cuatro me da cero, a dos le quito dos me da cero luego, ah! Y aquí ya mi respuesta siempre pero siempre va a ser el número el número de arriba y mi respuesta es el número siete..... el siete... y aquí no hay que contar porque la respuesta siempre va a ser la de arriba tonces así es como se hace la división”.</p>
<p>Ítem 12 Forma A</p>	<p>“Nueve dividido por setenta y dos... cuál me daría setenta y dos... nueve por seis... nueve por una... nueve por dos... dieciocho... nueve por tres... veinticuatro... nueve por cinco... treinta tres... nueve por cuatro... cuarenta y seis... nueve por cinco... cincuenta y cinco... nueve por seis... sesenta y cuatro... nueve por siete... nueve por siete!... aquí pongo el siete... llevo el siete acá... setenta y dos menos setenta y dos... cero... ahora lo sumo... sería esta respuesta... porque dio nueve por siete... sería setenta...aquí en la centena”.</p>
<p>Ítem 13 Formas A y B</p>	<p>“Esa si no me la han enseñado... Noo, esa si no me la han enseñado, de verdad?.. Aaaa, el diamante... Estoy pensando, 4”.</p> <p>“5... 7... 6”.</p> <p>“Hmm.... (dibuja un hexágono)... Nooo... Nooo... Un rombo... 6”.</p> <p>“6”.</p>
<p>Ítem 14 Forma A</p>	<p>“La número dos... porque el cuadrado tiene cuatro lados”.</p> <p>“El dos... porque esto es un rectángulo... porque el cuadrado tiene, sus medidas son exactas”.</p> <p>“La dos... porque los cuadrados tienen los cuatro lados iguales”.</p> <p>“Es la dos... porque es más chiquito que el primero”.</p>
<p>Ítem 14 Forma B</p>	<p>“El cuatro”.</p> <p>“Porque aquí hay un cuatro”.</p> <p>“La cuatro”.</p>
<p>Ítem 15 Forma A</p>	<p>“El círculo, el triángulo y el cuadrado... En el triángulo está el 2 y el 3; y en el cuadrado está el 4... (pausa larga).... Y escribo los tres resultados, o solooo... 1, 2, 3 y 4 están allí... A ver... Como es.... ¿Qué número es el que está en el... cual?... ¿Qué numero está en el triángulo y también en el cuadrado? Aaaaa, el 3.... Si.... Sí, porque no me acordaba y ahorita me acababa de acordar... De que estaba aquí en los dos”.</p> <p>“El 2 y el 3.... ¿Qué numero está en el triángulo y también en el cuadrado? En el cuadrado está el 4... 3... 1, 2, 3 y 4... El 4... Porque está en el cuadrado... Yyymmm.... En el triángulo está el 2... El 4... En el triángulo eeeee 3... Porque es en el triángulo... En el triángulo... en el cuadrado, 3 y 4. En el triángulo 2 y 3... 2, 3 en el triángulo y en el cuadrado 3 y 4... 2 y el 4”.</p> <p>“El 2 y el 3... en la rueda y en el cuadrado.... En el cuadrado esta el 4... Y en el triángulo está el 2... El 1, el 2, el 3 y el 4... Hmmm un círculo, un triángulo y un cuadrado... Hmmm... En el triángulo hay dos, dos números... El 4... En el triángulo pienso yo que es el 3... Porque es el triángulo y el cuadrado”.</p> <p>“El 3, porque está en la esquina del cuadrado y del triángulo...”</p>
<p>Ítem 15 Forma B</p>	<p>“El dos y el tres... no... el uno y el dos... el dos”.</p> <p>“El dos”.</p> <p>“El cua... pero de los números... el uno y el dos... en el círculo... el dos... porque aquí está el cuadrado, aquí está el círculo”.</p>

<p>Ítem 16 Forma A</p>	<p>“La número a)... hice como que sumé porque tres más tres, sería seis... y seis más ocho, catorce”.</p> <p>“Ya sé, es doce... porque aquí como este mide igual que este, pongo este número igual acá... y este acá... y como este, el tres más cuatro es seis, y como estos también, aquí también son números como este acá, aquí va un cuatro y un tres... y aquí es... y sumo estos dos seis y formaría doce”.</p> <p>“Sería la d... porque cuatro más tres es siete”.</p> <p>“Aquí tengo que sumar tres metros más cuatro metros, y me da siete, siete metros”.</p>
<p>Ítem 16 Forma B</p>	<p>“Emmm sería, diez... porque como estos dos lados miden tres metros, entonces serían tres más tres seis y como aquí mide dos metros, entonces serían dos más dos cuatro y sumo cuatro más seis que serían diez”.</p> <p>“Ah... Esta figura mide cinco metros de distancia... porque sumé dos.... ah no vamos a ver no, mide diez metros de distancia porque sumé dos más dos que mide estos lados y tres más tres que da diez”.</p> <p>“En sus cuatro lados, en este lado mide dos metros entonces el otro lado también sería dos metros ¿verdad?... en sus cuatros lados verdad, entonces sí... dos metros, cuatro, cuatro más tres ess, cuatro más tresss siete más tres sería once pero aquí no está la respuesta once, entoncesss.... En sus cuatro lados, entonces aquí tendría que poner los dos metros y si entonces allí ya sería cinco metros”.</p> <p>“Dos metros ¿es dos metros?... entonces mediría cinco metros?... si aquí mide dos entonces aquí mide dos metros también, aquí tres metros y aquí también tres metros, y aquí ya no sé cómo, mediría cinco metros... sólo sumé estos dos... dos metros y aquí tres metros”.</p> <p>“Aquí tengo que sumar los cuatro lados a mí me lo enseñaron emm como digamos que aquí hay otros números, esos números digamos que es dieciséis y quince catorce y trece tonces yo los tengo que sumar y ese número me va a dar la respuesta a mí... Tonces hay que sumar tres más dos, cinco centímetros”.</p>
<p>Ítem 17 Forma B</p>	<p>“Porque la balanza no puede medir el tiempo, el tiempo va a ser el reloj, la estatura tampoco puede ser porque uno va a ver eeel, hay ¿cómo se llama? El medidor y la temperatura lo mideeen, hay ya se me olvidó cómo se llaman... el termómetro... con el peso lo puede medir la balanza, el peso sí lo puede medir la balanza”.</p> <p>“Mm la balanza mide el peso... Lo he visto en la tele o en los mercados también”.</p> <p>“El peso, es la respuesta b... porque el tiempo no tiene nada que ver con la balanza, la estatura no importa y la temperatura sólo es airee, el peso sí porque si pongo un borrador y un lápiz la balanza se va a mover y va a decir cuál pesa más”.</p> <p>“Aquí sería el peso... porque aquí sería la misma cantidad pesando un objeto con otro objeto, porque el tiempo no puede ser porque si no pondría un reloj, la estatura no sería tampoco la temperatura, entonces sería el peso”.</p> <p>“El peso... porque digamos que aquí hay una balanza y hay que digamos que ponemos paquete de un paquete de arroz... y en el otro paquete ponemos un borrador, entonces la balanza baja o sube y lo que más pesa es el arroz y por eso es que se baja y el borrador sube”.</p>
<p>Ítem 18 Formas A y B</p>	<p>“Esta... porque tiene tres”.</p> <p>“En tercios son... tres... La que dice tres, cabal”.</p> <p>“La tres... porque tiene tres”.</p>
<p>Ítem 19</p>	<p>“La número tres... porque aquí tiene tres pintadas y aquí no tiene nada, tiene tres pero sin pintado”.</p>

Forma A	<p>“La tres... porque dice que sombree... cuál es las figuras está sombreada hasta la mitad. En la tres hay seis y está sombreada tres y en blanco tres”.</p> <p>“Sería la tres... en total tiene seis, tiene tres sombreadas y tres no”.</p> <p>“La tres... porque aquí hay seis y solo hay tres sombreadas y tres que no son sombreadas”.</p>
Ítem 20 Forma B	<p>“Dieciocho... antes. Dieciséis”.</p> <p>“El dieciséis”.</p> <p>“Dieciséis”.</p>
Ítem 22 Forma A	<p>“Sería el miércoles... porque dice que tiene trescientos ochenta y cinco, los demás trescientos veintiocho, trescientos setenta y dos y trescientos cuarenta y siete”.</p> <p>“Sería miércoles... Estaba quitando los demás números y miraba que son tres... todos tienen tres entonces me pasé al segundo número y miré que aquí habían dos, siete, ocho, cuatro... y vi que el ocho era mayor que los demás”.</p> <p>“El miércoles, que tiene trescientos ochenta y cinco, y las demás tienen un número menor de trescientos ochenta y cinco”.</p> <p>“Fue miércoles... porque vendió trescientos ochenta y cinco tortillas... porque los demás son menos que él”.</p> <p>“Va... Día, lunes, martes, miércoles y jueves. Cantidad: 328, 372, 385, 343. ¿Cuándo la cantidad de tortillas que vendió la señora?... Aaaaaaa (pausa larga)... Aaa, no aquí estoy viendo, cual es, que día fue?... el miércoles... Porque allí está, que el lunes está 328, en el martes 372, en el miércoles 385 y en el jueves está 347. Entonces me di cuenta que era en el miércoles”.</p> <p>“Luneeesssss, jueves!... El miércoles... Porque tiene el número más alto... c”.</p> <p>“Esteee días está: lunes, martes, miércoles y jueves... Y cantidad está 328, 372, 385 y 347... Es 385... la cantidad de 385 tortillas vendió... Y el día fue miércoles”.</p> <p>“El miércoles. Porque tiene... la cantidad es más que los otros números”.</p>
Ítem 22 Forma B	<p>“Depende de la cantidad, porque trescientos sesenta y nueve, es más grande que trescientos cincuenta y siete y trescientos ochenta y cuatro es más grande que todos los demás”.</p> <p>“El día en que la señora vendió más pan en los cuatro días fue el lunes... ah no, el miércoles porque lo analicé en mi mente”.</p> <p>“La respuesta es el miércoles porque aunque todos estén con el tres al principio pero el segundo dígito hay un seis, cinco, ocho y un cuatro y el miércoles es el que tiene el ocho entonces el que tiene más cantidad es el miércoles es la respuesta C”.</p> <p>“La respuesta sería el lunes porque vendió más pan, ahhh nooo vendió más el día miércoles con trescientos ochenta y cuatro panes”.</p> <p>“Mmmmm....el cuadro.....en cuatro días.....más pan mjm el miércoles... Porque..... el lunes.....el martes trescientos sesenta y nueve, miércoles trescientos setenta y siete, jueves trescientos cuarenta y ocho y el viernes doscientos setenti cuatro”.</p>
Ítem 23 Forma A	<p>“Son las 8, las 8 en punto... o sea que falta una hora”.</p> <p>“Las 8 en punto... una hora... Para las 9... A”.</p> <p>“8 en punto... 5 horas... Hmmm faltan... Las 8 en punto (escribe: 8:12)... Eeem 8”.</p> <p>“Las 12 y 40... Falta 9 horas, 8 horas”.</p>
Ítem 25 Forma B	<p>“Multiplicar... mmm... Restar... cien menos treinta... cero menos tres... el uno le presta al cero... diez menos tres... siete”.</p> <p>“¿Cómo así? ¿Qué número es igual a cien menos treinta? ¡Ah! O sea que cien menos treinta”.</p>

	<p>“Cien... pero no sé cuál es... ¿qué número es igual a cien menos treinta?... es ciento veinte... no”.</p>
<p>Ítem 26 Forma B</p>	<p>“Porque siete más tres es diez”.</p> <p>“El número que le tengo que sumar al siete para que me dé diez es tres”.</p> <p>“Es la respuesta d... siete más tres es diez”.</p> <p>“Sería la respuesta D, tres porque siete más tres es diez”.</p> <p>“Qué número le debo sumar a siete para que me de diez, tres números, la d”.</p>
<p>Ítem 29 Forma A</p>	<p>“Suma, setenta y cinco más cien... es la número c... porque hice cinco más cero, y siete más cero y el uno se quedó solo”.</p> <p>“Ciento setenta y cinco... porque puse este setenta y cinco en el cien”.</p> <p>“Fácil, ciento setenta y cinco... porque cien más setenta y cinco da igual a ciento setenta y cinco”.</p> <p>“Tengo que sumarlas, setenta y cinco más cien... me da... aquí sumo cinco más cero, y el cero no es nada así que bajo el cinco, y siete más cero que bajo el siete y aquí solo bajo el uno porque está solito... y compró ciento setenta y cinco frutas”.</p>
<p>Ítem 29 Forma B</p>	<p>“Sumar... cien más ochenta y cinco... ciento ochenta y cinco”.</p> <p>“Cien más ochenta y cinco. Porque compró cien duraznos y ochenta y cinco peras; y la pregunta es ¿cuántas frutas compró en total? Entonces se suman cien más ochenta y cinco y son ciento ochenta y cinco”.</p> <p>“¿Pero aquí tengo que sumar?... cien más ochenta y cinco... sumar cien más ochenta y cinco... aquí cinco... cero más cinco... Cinco... porque el cero no tiene, no tiene nada... cero más ocho, ocho... queda el uno... porque no hay ningún número”.</p> <p>“Mmm, como los duraznos son frutas y las peras también como los dos son frutas sumo cien más ochenta y cinco y da ciento ochenta y cinco”.</p> <p>“Aida en total compró 185 frutas... porque sumé cien más ochenta y cinco y da ciento ochenta y cinco”.</p> <p>“Ya sé la respuesta pero dice que tengo que hacer esto, entoncesss cien más ochenta y cinco vaa, como es una suma cero más cinco es cinco cero más ocho es ocho y uno que quedó sólo bajamos el uno, es ciento ochenta y cinco y la respuesta es C”.</p> <p>“Sería ciento ochenta y cinco... porque Aída compró cien duraznos y ochenta y cinco peras ¿cuántas frutas compró? Entonces van cien frutas y ochenta y cinco entonces serían ciento ochenta y cinco... una suma”.</p> <p>“Mmm mmm ay allí si está un poquito, Ah! Ciento ochenta y cinco... Porque bueno yo lo hago en por ejemplo venía así en el examen entonces pero yo en un papel cuando no se daba cuenta la maestra escribía que me cuesta ah entonces ponía cien emm sumad, sumado en ochenta y cinco bueno hay dos ceros y el cinco hay que bajarlo y el ocho también y el uno también entonces me dio ciento ochenta y cinco”.</p>
<p>Ítem 30 Forma A</p>	<p>“Resta... ¡no, suma!... cien diez más ciento cinco y en total sería doscientos quince... porque sumé cinco más cero, cero más uno y uno más uno”.</p> <p>“Doscientos quince... porque sumé uno más uno, es dos... y como aquí el uno más cero lo pongo aquí en el uno, y aquí el cinco... junté los dos números”.</p> <p>“Doscientos quince... porque ciento diez más ciento cinco me da doscientos quince”.</p> <p>“Aquí tengo que sumar ciento diez más ciento cinco... aquí solo bajo el cinco porque es un cero, aquí también solo bajo el uno, porque también tiene un cero... y solo sumo uno más uno que me da dos... y la respuesta es doscientos quince”.</p>

<p>Ítem 30 Forma B</p>	<p>“Sumar... ciento cinco más ciento veinte... mmm cero más cinco... cinco... dos más cero. Dos... uno más uno. Dos... doscientos veinticinco”.</p> <p>“Igual, ciento cinco más ciento veinte... cinco más cero, cinco. Dos más cero, dos. Y uno más uno son dos”.</p> <p>“Sumar... ciento cinco más ciento veinte... cinco más cero... cero... cinco... cero más dos, dos... uno más uno, dos... doscientos veinticinco”.</p>
<p>Ítem 31 Forma A</p>	<p>“El nueve menos tres sería, y cuatro menos cinco no se puede y a la par había un uno... entonces le hice como un... se quedó en valor de cero y el otro se quedó en valor de catorce... hice catorce menos cinco”.</p> <p>“Voy a restar tres menos nueve, serían seis... aquí no lo puedo hacer porque este es número menor que este... así que voy a poner este, queda en cero y este se convierte en catorce... y quedaría, y este quedaría como un diez... aaah no, quedaría como un nueve... entonces sería noventa y seis”.</p> <p>“Tendría que restar ciento cuarenta y nueve menos cincuenta y tres... entonces a nueve le quito tres y seis... a cuatro le quito cinco, no se podría... entonces el uno le presta uno al cuatro, entonces serían catorce menos cinco que me da igual a once... entonces... No, no, no, no... me daría diez... no... nueve, nueve... entonces sería noventa y seis... subrayo la respuesta a que sería la a”.</p> <p>“Aquí tengo que restar ciento cuarenta y nueve menos cincuenta y tres... aquí nueve menos tres... me quedan seis... cuatro menos cinco no se puede, entonces le pido prestado uno al uno y me da catorce, a catorce le quito cinco... me quedan... me quedan nueve... y el resultado es noventa y seis”.</p> <p>“Va, ciento, cuarenta y nueve va... a 9 le restamos 3, nos queda 6. A 4 le restamos 5, nos e puede entonces el 1 le presta al 4 y este se convierte en 14. Va, ahora a 14 le quitamos 5, quedan 9 y el 1 quedo en valor de 0. Entonces 96, la a”.</p> <p>“53 más 149; 149 más 53... No, quitando... 149 y le quito 53... Tengo 9 y le quito 3 = 6. (pausa) Tengo cuatro, 5 y le quito 4, porque con 4 no se puede... Quitarle al 5... No?... Como el 4 es más... Más pequeño que el 5, le falta un número para 5... El 9 le presta al 4... El 1 le presta al 4... 5 le quito 5 = 0... Y de allí el quedo en 0 (refiriéndose al préstamo del 1 al 4)... En 5”.</p> <p>“Tengo que dibujar 149 cajas para saber cuántas cajas quedaron... Borrar 53 cajas... Hmmmm resta... Hmmmm... Aquí no se puede... Porque arriba esta el 1 y el 5 tiene que estar arriba, no el uno”.</p> <p>“Tonces 149 menos 53... entonces: 9 menos 3 es 6; 4 menos 5 no se puede entonces le quitas un 1 al 100 y te queda 14, entonces 14. 14 menos 5 es 9, 9... Y cero menos nada es 96, 0. Entonces es 96”.</p>
<p>Ítem 31 Forma B</p>	<p>“Eemmm, voy a restar veinticinco menos cuarenta... entonces a cien le quito quince y quedan, aquí no dice vaa... Me quedan ochenta y cinco, comoo, le quito veinticinco a cuarenta quedan quince entonces a cien le quito quince y quedan ochenta y cinco”.</p> <p>“Ah, a Leonor le quedan ochenta y cinco chocolates... La resta”.</p> <p>“Ciento veinticinco que es el total de chocolates y cuarenta menos porque le regaló a su amiga, entonces cinco menos cero es cinco, dos menos cuatro no se puede entonces el uno se queda en cero y le presta uno al dos y se vuelve doce, se resta aquí y sería ocho y como el uno se quedó en cero es cero, entonces es ochenta y cinco y es la respuesta c”.</p> <p>“Entonces sería si ciento veinticinco menos cuarenta, porque Leonor le regaló cuarenta a su amiga, entonces sería operación de resta, ehhhh, cinco menos cero no se puede, entonces, ownn sí se puede... Entonces sería cinco menos cero aquí pongo</p>

	<p>el cinco, dos menos cuatro serían dos y éste número como no hay entonces, ahhh no este sería... cinco, dos menos cuatro no se puede entonces le quito uno y me queda en cero, dos le quedan doce entonces doce menos cuatro serian seriaannn ocho, sería me quedaría ochenta y cinco, entonces sería la C”.</p> <p>“Mmm ciento veinticinco más cuarenta, aquí puedo hacer mi... mmm aquí no era así mjm, mjm, mjm, aquí bueno a mí yo sumé ciento veinticinco por cuarenta más cuarenta y el cinco lo bajé y a cuatro le sumé dos me dio seis me dio la respuesta sesenta y cinco pero como había arriba un uno lo bajé pero acá creo que falta un número o estoy yo equivocada que yo creo que es aquí debería ir un uno, la respuesta sería ciento sesenta y cinco”.</p>
Ítem 32 Forma A	<p>“Va, 8... Porque hay 80 panes, va y poner 10 pones en cada bolsa”.</p> <p>“Aquí poner lassss, el número y las bolsas... 80... 10. Diez le quito... Suma... 80... y dieeee... y va a poner 10 panes en cada bolsa... nuev, 90!.. El ochoooo le presta 1 al 0”.</p> <p>“Hacer una suma... Para ver cuántas bolsas necesita para meter 10 panes en cada bolsa... (escribe $80 + 10$)... Una suma... 80 más 10... 80 más 10... : 9... Eee 9... la d”.</p> <p>“Entonces 10 dividido 80 es 8. Dividí porque dice 80 panes entonces: ¿Cuántas bolsas necesita si pone 10 panes en cada bolsa?... Entonces 8 por ochen, 8 por 1 es 1 y 8 por 0 es 0”.</p>
Ítem 33 Forma A	<p>“Cuatro... porque tres más uno son cuatro... en donde dice la pregunta: estuvimos tres semanas en la playa y una semana en Antigua”.</p> <p>“Allí es suma... tengo que sumar el tres con el uno... tres más uno son cuatro... porque allí dice estuvimos tres semanas en la playa una semana en Antigua... son cuatro viajes”.</p> <p>“Cuatro... porque tres más uno... cuatro... tres semanas más una...son cuatro”.</p> <p>“Siete... catorce...quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte, veintiuno... más uno... veintidós... no está la respuesta... solo veintiuno. Allí sumé... las tres semanas con un día... las tres semanas las sumé... Ahhh una semana... Veintiuno, veintidós, veintitrés, veinticuatro, veinticinco, veintiséis, veintisiete, veintiocho... veintiocho días estuvieron de tres semanas y una semana... serían cuatro... tres semanas más una semana... estuvieron veintiocho días”.</p>
Ítem 34 Forma B	<p>“Mmmm, como Jorge pagó con un billete de a cien serían doscientos entonces a ciento setenta y cinco le sumoo por ejemplo veinte y me dan ciento noventa y cinco, nooo, hay nooo, le sumo veinticinco entonces sería cinco más cinco diez llevo uno aquí y entonces dos más siete son nueve más uno son diez, son diez entonces allí sé que son veinticinco”.</p> <p>“Aquí mm resté para ver cuánto le tienen que dar de vuelto a Jorge y así es correcto... Resta... emm doscientos menos ciento setenta y cinco y me tienen que dar veinticinco quetzales de vuelto”.</p> <p>“El vuelto sería de veinticinco... hago la operación?... estoy poniendo lo que él pagó, que es cien más cien doscientos y aquí a él, voy a poner doscientos menos ciento setenta y cinco, entonces allí me va a decir cuánto de vuelto le tienen que dar. Éste le presta uno al cero y éste se queda en valor de nueve y éste en valor de diez porque no le presta a nadie, diez menos cinco da cinco, nueve menos siete es tres, es dos perdón y uno menos uno es cero es veinticinco vaaa”.</p> <p>“Eso es una resta... Sería doscientos menos ciento setenta y cinco... eh hh, porque Jorge compró un juguete que le costó ciento setenta y cinco, pagó con dos billetes de</p>

	<p>cien ¿cuánto le dan de vuelto? Tengo que restar... serían doscientos menos ciento setenta y cinco, serían cero menos cinco no se pueden y tenemos que prestar, el dos queda en uno y éste cero se queda en nueve y éste se queda en diez porque le prestó uno a éste, entonces sería diez menos cinco da cinco y nueve menos siete serían dos y uno menos uno serían cero, le dieron veinticinco quetzales de vuelto sería respuesta B”.</p> <p>“Aquí también tengo que hacer una aquí te la voy a poner arriba... Aja aquí, lo tengo que hacer por ejemplo tengo que hacer todo este por todo esto tengo que agre, tengo que agregarle los dos ceros o solo así como lo puse yo... mjm, muy bien aquí lo sumé y como te expliqué que aquí los ceros no se cuentan tonces este bajé el cinco bajé el siete y sumé uno más uno entonces mi respuesta sería doscientos setenta y cinco, ésta respuesta sería”.</p>
<p>Ítem 35 Forma A</p>	<p>“Sumar...porque decía... aquí cuántos... treinta y nueve más ocho”.</p> <p>“Allí creo que es suma... aquí bajo el tres... y aquí son quince... hay por todos... trescientos quince”.</p> <p>“Cuarenta y ocho canastas... treinta y nueve más ocho son cuarenta y ocho”.</p> <p>“Aquí en treinta y nueve... mmm... aquí sería treinta y nueve por ocho... daría... Borro porque me equivoqué de signo... aquí sería ocho por nueve... ocho por nueve noventa... ochenta y nueve... ochenta y ocho... ochenta y siete... ochenta y seis... ochenta y cinco... ochenta y cuatro... ochenta y tres... ochenta y tres... pongo ocho y llevamos tres... ocho por tres... ocho... dieciséis, dieciocho, diecinueve, veinte, veintinueve, veintidós, veintitrés, veinticuatro aquí más tres... veinticinco, veintiséis... aquí pongo el seis y el dos... doscientos sesenta y ocho. Entre todos fueron doscientos sesenta y ocho... multipliqué ocho por nueve y ocho por tres... porque aquí dieron treinta y nueve por ocho... lo tuve que multiplicar aquí”.</p>
<p>Ítem 36 Forma B</p>	<p>“Multiplicar... tres por treinta... sumar... diez más dos... treinta por tres... tres por cero, cero. Tres por tres seis, nueve... elotes... diez... dos... dos canastos con diez tomates... doce... diez por dos... porque hay dos canastos con diez tomates... dos por cero, cero. Dos por una, dos... ¿cuántas verduras tengo en total?... sumar diez más veinte. Noventa más veinte... cero más cero, cero. Once... ciento diez”.</p> <p>“Si son tres canastas con treinta elotes son... noventa elotes. Y si son dos canastas con diez tomates son veinte tomates. Entonces, ¿cuántas verduras tengo en total? Son noventa más veinte... cero, once. Ciento diez”.</p> <p>“Sumar o res... sumar... mmm se puede de primero treinta más diez... tres... treinta... tres... treinta... tres... treinta más tres... treinta... es treinta... pero ¿cuál sumo?... diez más treinta... aquí cero... cuatro... cuarenta... y restando... diez más treinta, aquí igual cero... pero aquí no se puede... porque aquí solo es uno y aquí tiene más... multiplicando... diez por treinta... multiplicando cero por cero... cero... uno por tres... tres... treinta... y treinta... tres por dos... treinta por tres... tres... cero... tres por tres... seis... nueve... diez por dos... dos por cero, dos... cero... uno por dos, dos... que hay veinte tomates... aquí... ¿cuántas verduras tengo en total?... canastos... noventa... ¿de tomates y de elotes?... sumar estos dos... noventa más veinte... aquí cero; nueve más dos, once... ciento diez”.</p>
<p>Ítem 37 Forma A</p>	<p>“O sea que si tenía quince niños y setenta y nueve chocolates... es resta... debería de ser setenta y nueve menos quince”.</p> <p>“Dividir... setenta y nueve dividido quince... me di cuenta de que, estuve sumando quince, más quince, más quince hasta llegar a cinco que sería setenta y cinco, porque si sumarlo una vez más llegaría a ochenta y cinco... en esta parte yo lo voy a, voy a</p>

	<p>restarlo... cuatro me quedaría”.</p> <p>“Sobrarían... tendría que restar... tendría que... setenta y nueve menos quince... si a nueve le quito cinco me daría... multiplicar... ah no, dividir... entonces divido setenta y nueve dividido quince... un número multiplicado... quince... no me recuero mucho qué tenía que hacer en una división”.</p> <p>“Hay que dividir... y la respuesta es tres”.</p>
<p>Ítem 37</p> <p>Forma B</p>	<p>“Hay que restar... ¿multiplicar?, ¿sumar?, ¿restar?, dividir... cincuenta y uno divido doce... cinco dividido uno... uno... uno dividido dos... dos... no”.</p> <p>“Cincuenta y uno menos doce. Como no se puede son once. Once menos dos son nueve y como este quedo en cuatro, cuatro menos uno son nueve. Treinta y nueve. ¿Por qué?... ¡ah! Entendí mal. No se puede... entonces es división... son doce, cincuenta y uno ¿verdad? ¡eh! Doce por dos son veinte y cuatro; por tres son treinta y seis; por cuatro son cuarenta y ocho y por... si son cuarenta y ocho porque de ahí son mucho. Entonces son por cuatro. Once menos ocho son... tres y cuatro menos uno trece y uno... son once y sobraron tres... cinco menos uno... cuatro menos uno tres”.</p> <p>“Mmm sumar... restar... ¿multiplicar?... poniendo unos palitos... para sumar... si... mmm restar, tengo que poner doce... si... diez, once, doce... treinta y siete si reparto tres... si... treinta y ocho, treinta y nueve, cuarenta, cuarenta y uno, cuarenta y dos, cuarenta y tres, cuarenta y cuatro, cuarenta y cinco, cuarenta y seis, cuarenta y siete, cuarenta y ocho, cuarenta y nueve... uno... si... si les reparto cuatro a cada niño, un, dos tres... etc.... cuarenta y cinco, cuarenta y seis!... cuatro”.</p> <p>“Voy a restarle, a nooo, no no, voy a dividir cincuenta y uno dentro de doce... para saber cuántos pastelitos le sobró a Daniela... mjmmm, entonces tengo que multiplicar doce pooor por cinco, noo, siiii, me equivoqué otra vez.....entonces aquí no puede ser cincooo porque doce por cinco da sesenta , si da sesenta entonces tiene que ser un número menos... mjmmm, porque como aquí es cincuenta y uno y tiene que ser menor que cincuenta y uno y no mayor... multiplicar un doce por cuatro... emmm, ahhh, me dio cuarenta y ocho... como dos por cuatro es ocho y como siempre que tenés que multiplicar de dos dígitos tenés que correr uno, poner un cero aquí y aquí ponés la respuesta de éste por este entonces ocho más cero da ocho y como aquí no hay más pones cuatro y sería cuarenta y ocho, entonces aquí pongo cuatro pongo un cuarenta y ocho , aquí tengo un cuatro, éste se convierte en once y ocho menos once serían tres y cuatro menos cuatro es cero entonces ...sería... no, es emmm, mjmmmm... es que como es cuatro dividido tres tiene que ser uno de éstos dos y yo digo que es el tres... porque es el residuo o sea lo que queda... entonces es tres , pero si me preguntaran cuánto le da a cada niño serían cuatro y no tres... tres”.</p> <p>“Porque le tiene que repartir a doce niños la misma cantidad a cada uno de pastelitos... mmm el cincuenta y uno tiene que ser afuera porque es lo que vamos a dividir y el doce adentro..... Mmm aquí la respuesta es cinco y nos sobraron seis... porque mm ese mm creo que es el que da la respuesta correcta... no es que me confundí en esta... esta si no la entiendo... mejor la voy a dejar en blanco porque no la entiendo”.</p> <p>“Mjjmmmmmm, “entreee”... lo estoy pensandooo...entonces pudiera repartir, podría hacerlooo en doce dividirlo entre doce, pero lo hice muy arriba vaaa...doce por sieteeee tampoco, entonces doce por cuatrooo es cuarenta y ocho que es la respuesta, lo restooo uno menos ocho no se puede entonces le quito uno al cinco y el uno se vuelve en once , once menos ocho es trece y cuatro menos cuatro es cero , la respuesta es la a , que es tres... la respuesta ess.... Los tres no los dio entonces ella se va a quedar con ellos, porque tres es como lo que sobró”.</p>

	<p>“Aquí sería una resta, sería cincuenta y uno menos doce, como el uno no se puede restar entre dos reagrupamos, entonces el cinco se queda en cuatro y el uno se queda en diez, noo se queda en once, ehheh once menos dos sería nueve, menos dos no se puede entonces se queda en cuatro y esto se queda en cuatro, once menos dos es nueve y cuatro menos uno, no see entonces sí estaría malo... porque ella repartió cincuenta y un pastelitos entre doce niños, tengo que restar cincuenta y uno menos doce... porque no se puede restar... entonces tengo que hacer una división... para dividir doce entre cincuenta y uno... doce en cinco no cabe entonces, doce por una doce doce, doce por dos veinticuatro, doce por tres cincuenta y seis, sería doce por cuatro, aquí serían serían cincuenta . Sería uno menos cero uno y cinco menos cinco cero...”</p> <p>“Repartió cincuenta y uno entre doce niños aquí sería una resta... Porque aquí dice Daniela repartió cincuenta y uno pastelitos... entre doce niños, le dio la misma cantidad a cada uno cuantos pastelitos le sobraron, será una suma o será una resta... no es una suma entonces a dos le resto uno queda uno, aquí entonces me queda cuatro entonces mi respuesta creo que sería mmm..... Ah! Aquí bueno mi respuesta sería cuarenta y uno”.</p>
<p>Ítem 38 Forma A</p>	<p>“A veeer, ochooooo... Multiplicar!... 8 por 9... Bueno eso es lo que pienso yo, pienso no sé si es cierto... A ver 8 la hora (pausa larga)... estoy pensando... Va, 9 por 7... Va, serian, 9 por 8 o 9 por 7...?.../ 9 horas vamos a ver, 9 horas empezamos. Va, a 8 le quitamos los 3 entonces lo dejamos en 5... Vamos a ver, a ver empezamos.... 5, 10, 15, 20, 25, 30”....</p> <p>“Es suma... Porque le dan dinero... 9 más 7 y 7 más 8... Sumar... 9 más 7... 16 horas... Por trabajar 8 por hora, le pagaran por los dos días de trabajo 126”.</p> <p>“Aquiiii... sumar, sumar cuantoos quetzales le pagaron a Da... Eeee... el 9, el 7 y el 8... Hmmm sumar... 9 más 7... Mmmmm, mmmm... es ... Mmmmm... 17... Hmmm no”.</p> <p>“Entonces 9 por 7... Porque el, estos números fueron las horas que él trabajó. Entonces lo multipliqué... Porque estas dos horas hacen un número que es 63 y lo puedo sumar por 800; porque son 9 centavos por 7 centavos, 63 centavos más 808 quetzales, hace 868 quetzales con 63 quetzales... Sumando (con voz dudosa) 9 más 7... Porque... multiplicar no funcionó, así que mejor eee.. sumarlo... 9 más 7 es 16... Entoncess... maaas.... Lo que multiplique aquí... 9 por 7 que da 63”.</p>
<p>Ítem 38 Forma B</p>	<p>“Tres mjmmmm aquí serían doce por cinco serían sesenta que es una hora eeemm... como dice que tarda doce minutos para caminar un kilómetro entonces multiplicamos doce por cinco que serían sesenta y eso es una hora entonces ahora doce por cuatro que serían cuarenta y ocho pero, si juntamos sesenta más cuarenta y ocho serían , noo tiempo, doce por cinco serían sesenta y doce por cuatro serían cuarenta y ocho... ajá, ciento ocho entonces Margarita caminó ciento ocho minutos en dos días... como doce por cinco es sesenta y sesenta es una hora , como cuatro por doce es cuarenta y ocho , entonces como sesenta más cuarenta y ocho dan ciento y ocho”.</p> <p>“Mmm porque cuantos minutos caminó en los dos días Margarita ... mmm y se tardó ciento ocho minutos en dos días... emm multipliqué para ver si me salía una respuesta de estas para ver cuantos minutos había caminado Margarita... emm primero multipliqué cinco por doce y me salió sesenta y después multipliqué cuatro por doce y me salió cuarenta y ocho y después mmm sumé sesenta más cuarenta y ocho y me salió ciento ocho”.</p> <p>“Entonces cuatro kilómetros del sábado y cinco del domingo si tardara doce minutos</p>

	<p>para caminar un kilómetro entonces aquí tendrían que ser dos operaciones, o no? Veamos... entonces vamos a poner nueve para que se me haga más fácil y doce que es lo que se tarda, lo voy a multiplicar por doce, es que esto lo hice muy arriba por doce, allí está.... Doce por nueve, dos por nueve es ocho y uno porque es dieciocho y el uno aquí arriba, nueve por uno es nueve más uno es diez es ciento ocho y ciento ocho está en la respuesta C, es C, ya termineee... el nueve lo saqué porque cinco más cuatro es nueve y para no hacer dos operaciones lo sumé y lo puse con nueve y el doce porque cada vez tarda doce minutos , entonces se tardó ciento ocho minutos , eso se tardó en los dos días, eso se tardó”.</p> <p>“Aquí haría también una división...a nooo, aquí tengo una multiplicación... tendría que sumar... mmm, no se puede sumar porque si no sumaría mucho... una división... no see... eso estoy analizando....aquí tendría que hacer unaaa, unaaa multiplicación, no estoy seguro... está mala... no estoy seguroooo... siiii..... es una división?... es que no sé cómo se hace una multiplicación así... no me aparecía el signo, no sé cómo hacerlaaa... doce por nueve... doce y abajo sería nueve, nueve por dos dieciocho , pongo el ocho y nueve por una nueve más uno diez , ciento ocho sería la respuesta... de ciento ocho minutos”.</p> <p>“Aquí pongo cinco pongo doce con cuatro entonces hay que sumarlo y dice aquí puedo hacerlo de otra forma no necesariamente como está aquí puedes poner a veces doce cinco y cuatro entonces aquí pres pues sumo doce más cinco; trece, catorce, quince, dieciséis, diecisiete más cuatro: dieciocho, diecinueve, veinte, veintiuno aquí dio veintiuno tonces no sé si la respuesta está mal o las que están ahí... Y yo calculo que es veintiuno... Ah estos números como aquí dice Margarita caminó cinco kilómetros el sábado tonces aquí puse el cinco porque yo sumo los números tonces y cuatro kilómetros el domingo pongo el cuatro o el doce o así como tarda doce minutos pongo el doce y lo sumo para saber cuánto tonces yo calculo que veintiuno kilómetros a veces lo escribo a la par ahí está”.</p>
<p>Ítem 39 Forma A</p>	<p>“Restando... porque Julio estaba cosechando ochenta y ocho libras de maíz y alguien vino y le pidió veinticinco libras”.</p> <p>“Restar... porque dijo que vendió, o sea eso dejó de hacerlo... le vendió para dejar de vender... vendió sus 25 libras y tuvo que restar”.</p> <p>“Dividiendo... restando... resta...”</p> <p>“Restando... porque se vendió”.</p>
<p>Ítem 40 Forma A</p>	<p>“Va Yolanda lo armó más despacio que Rodrigo, pero más rápido que Arturo. ¿Quién lo armó más rápido de todos?... Rodrigo... Porque yo digo... Aquí dice que Rodrigo la armó más rápido que todos. Porque Rodrigo lo armó más rápido que Daniel y Arturo lo armó más despacio... Y como lo armo más rápido que Yolanda, Rodrigo fue quien termino más rápido”.</p> <p>“Rodrigo, y Arturo, Daniel y Yolanda están armando rompecabezas. Rodrigo y Arturo lo armaron más rápido... queee Daniel y Yolanda... Yolanda lo armó más despacio que Rodrigo, pero más rápido que Arturo. ¿Quién lo armó más rápido de todos? Yolanda! ... Porque lo armó más despacio... Porque lo armó más despacio que Rodrigo, Arturo y Daniel... ¿Quién lo armó más rápido que todos? Yolanda”.</p> <p>“Eeee fue Arturo... Fue Arturo y Rodrigo?... Hmm... porque ellos lo armaron más rápido... Que Daniel y Yolanda... Fueee... (pausa)... Fue Rodrigo... Porque él lo armó más rápido... Porque aquí diceee: Yolanda lo armo más despacio que Rodrigo, pero más rápido que Arturo... La respuesta fue Rodrigo”.</p> <p>“Rodrigo, porque dice que Yolanda lo armó más despacio que Rodrigo pero más rápido que Arturo. Entonces Rodrigo lo armó más rápido”.</p>

Ítem 40 Forma B	<p>“El lugar donde compró la camisa. El color de la camisa. El estilo de la camisa. El precio de la camisa... el precio de la camisa... porque él quería saber cuánto le darían de vuelto”.</p> <p>“Se compró una camisa y pagó con un billete de a cincuenta. Veamos... el lugar donde compró la camisa, no; el color... ¿aquí solo hay que subrayar?... El color de la camisa. El estilo de la camisa. ¡El precio de la camisa!... porque se le resta y sabe uno cuánto es de vuelto”.</p> <p>“El lugar donde compró la camisa. El color de la camisa. El estilo de la camisa. El precio de la camisa... el precio de la camisa... para ver cuánto le dan de vuelto”.</p>
--------------------	---