

# Validez de constructo del Sistema Cualitativo de Calificación para el Test Gestáltico de Bender Modificado

César Merino Soto<sup>1</sup>

---

<sup>1</sup> Instituto de Investigación en Psicología,  
Universidad de San Martín de Porres, Lima

---

Perú

*Correspondencia:* César Merino Soto. Dirección postal: calle Filiberto Romero 430, Chorrillos, Lima 9, Perú. E-mail: [sikayax@yahoo.com.ar](mailto:sikayax@yahoo.com.ar)

---

© Education & Psychology I+D+i and Editorial EOS (Spain)

## Resumen

**Introducción.** Se presenta un estudio entre un nuevo sistema de calificación (Sistema de Calificación Cualitativa) para la versión modificada del Test Gestáltico Visomotor de Bender Modificado (TGB-M) y una medida de habilidades pre-académicas en niños preescolares. En las investigaciones hispanas, la validez de constructo generalmente ha sido conducida con sistemas de calificación antiguos pero no con recientes propuestas.

**Método.** La muestra fue 70 preescolares entre 5 y 6 años de edad, en una zona urbana de Lima. Se aplicó el Sistema de Calificación Cualitativa al TGB-M. Para el análisis se normalizaron las puntuaciones y se corrigieron las correlaciones por atenuación del error de medición; finalmente, se aplicó un método correlacional multivariado (correlación canónica).

**Resultados.** Se hallaron correlaciones moderadas con las tareas de percepción visual y visomotricidad de la prueba de habilidades pre-académicas, mientras que relaciones nulas entre comprensión verbal y cálculo. Estas tareas fueron las únicas medidas psicométricamente más consistentes y sugieren una limitada validez de la prueba de habilidades pre-académicas.

**Discusión y conclusiones.** Se discuten estos resultados en el contexto de la utilidad de la nueva versión del TGB y del proceso de construcción de pruebas, específicamente la evaluación de en niños que ingresan a estudiar al primer grado elemental

**Palabras Clave:** validez, niños, visomotor, evaluación, Sistema de Calificación Cualitativa, Test de Bender

*Recibido: 04/10/11*

*Aceptación inicial: 05/10/11*

*Aceptación final: 02/11/11*

# Construct validity of the Qualitative Grading System for the Modified Bender Gestalt Test

## Abstract

**Introduction.** The aim of this study is the association between a new scoring system (Qualitative Scoring System) for the new Bender Gestalt Test Bender Modified (TGB-M) and a measure of pre-academic skills in preschool children. In the Hispanic research, construct validity has generally been conducted with older rating systems but not with recent proposals.

**Method.** The sample was 70 preschool children between 5 and 6 years-old, from an urban area of Lima. We used the Qualitative Scoring System for the TGB-M. For the analysis, the scores were normalized and corrected for attenuation of measurement error; and finally apply a canonical correlation.

**Results.** Moderate correlations were found with the tasks of visual perception and visuo-motor from the pre-academic skills test, while null relationships between verbal comprehension and calculation task. These tasks were the only psychometrically measures more consistent and suggest a limited validity of the test of pre-academic skills.

**Discussion and conclusion.** These results are discussed in the context of the usefulness of the new version of TGB and test construction process, specifically for assessment process for enter to elemental first-grade.

**Keywords:** validity, children, visuo-motor, assessment, Qualitative Scoring System, Bender's Test.

*Received: 10/04/11*

*Initial acceptance: 10/05/11*

*Final acceptance: 11/02/11*

## Introducción

El Test Gestáltico Visomotor de Bender (TGB) no necesita mucha presentación ya que su gran longevidad se ha mantenido, en parte, por los sistemas de calificación creados para cuantificar el desempeño medido por el tipo de tarea visomotora que requiere; es decir, el copiado de diseños geométricos. Se ha descrito al TGB como una medida de integración visomotora (Koppitz, 1984) o de habilidades visoconstructivas (Lacks, 1999), y el nivel de desempeño obtenido se lo ha asociado con déficits en diferentes áreas del comportamiento, como problemas de lectura, problemas emocionales y detección de problemas neurológicos (Koppitz, 1984; Cummings, Hoida, Machek y Nelson, 2003; Piotrowski, 1995; Perticone, 1998; Reichenberg & Raphael, 1992).

El TGB ha tenido varias modificaciones y una de ellas fue una versión abreviada, en que se usaron 6 láminas consideradas como las más apropiadas y discriminativas para niños escolares entre 4 y 8 años. Esta versión, así como su sistema de puntuación, proviene del trabajo de Keogh y Smith (1961) y de Hirsh (de Hirsh et al., 1966; Jansky y de Hirsh, 1972); esto últimos trabajaron directamente con L. Bender para elaborar un sistema de puntuación simplificado que permitiera la evaluación global de los diseños reproducidos por los niños. La versión abreviada de seis láminas del TGB, en sus inicios, se calificó con el Sistema Koppitz (1984), dentro de baterías de despistaje (screening) del riesgo de problemas académicas en los primeros años de escolaridad, demostrando que una parte de la varianza en el rendimiento de pruebas estandarizadas escolares puede ser explicado por esta versión (Telegdy, 1974a; Telegdy, 1974b; Telegdy, 1975; Wallbrown, Engin, Wallbrown y Blaha, 1975; Wallbrown, Wallbrown y Engin, 1974)

El contexto de la presente investigación es la evaluación de los correlatos de validez del Test Gestáltico Visomotor de Bender (Bender, 1938), pero específicamente con una reciente y apropiada en niños de los primeros ciclos de la educación primaria (Brannigan y Brunner, 2002). En su aplicación para niños, la versión que ha permanecido muy popular a través de los años y ha generado más de 300 artículos publicados (Reynolds, 2007) es el Sistema Evolutivo de Koppitz (SEK, Koppitz, 1984). Sin embargo, su eficacia ha sido variable respecto a su validez predictiva; algunos estudios han reportado que las correlaciones del desempeño en el TGB usando este sistema con criterios relevantes han sido de menor magnitud (Johnston y Lanak, 1985) que otros sistemas de calificación aún menos conocidos, por ejem-

plo el Sistema Watkins (1976). Más recientemente, se ha reportado que el Sistema Koppitz tiende a explicar menos varianza respecto a criterios de rendimiento escolar estandarizado que otros sistemas más recientes (Chan, 2000; Brannigan y Brunner, 1989, 1996, 2004; Parsons y Weinberg, 1993). Por lo tanto, apuntar el interés hacia nuevos enfoques y con mejor y actual respaldo científico es necesario para una apropiada práctica profesional. Pero las pruebas señaladas por sus autores como que evalúan atributos igualmente etiquetados por otras pruebas no deben asumirse como sustitutos intercambiables, porque cada uno puede demandar diferente tiempo para calificar los protocolos, un enfoque diferente de abordaje de la calificación y los niveles de acuerdo inter-calificador pueden ser bajos (Preda, 1997); y por lo tanto, producir divergencias en su validez de constructo cuando se los usan para evaluar los cambios en programas remediales y el desarrollo visomotor en general (Palisano y Dichter, 1989).

Hay numerosa investigación sobre los correlatos en el desempeño del Test de Bender, específicamente para el Sistema de Koppitz, pero estos provienen casi totalmente del mundo anglosajón. Por otro lado, una revisión informal de las investigaciones no publicadas y publicadas en el mundo hispano, a nivel de pregrado y postgrado, llevaría a resaltar que la evaluación de las propiedades psicométricas no es uno de sus objetivos principales al usar sistemas de calificación para el TGB. Por lo tanto, es muy infrecuente hallar resultados sobre el grado de acuerdo entre calificadores en estas investigaciones, o referencias directas a la comparación de coeficientes de validez entre sistemas de calificación, o sobre los correlatos con otros constructos. Los estándares propuestos por la asociación de AERA, APA y NCME. (1999) recomiendan que se reporten las evidencias de validez y confiabilidad para los usos propuestos del instrumento; y considerando que el TGB la evaluación de la integración visomotora continúa siendo un componente predictivo del rendimiento académico (Grupo ALBORCOHS, 2005; Simner, 1991; Simner y Barnes, 1991). Hay un número que baterías predictivas que los han incluido como una habilidad que mejore la predicción del aprendizaje durante los primeros años escolares, específicamente, en el inicio del primer grado (Berdicewski y Milićic, 2004).

Hasta la fecha, parece que las versiones del TGB diferentes a la versión desarrollada por Koppitz no han atraído tanto la atención de los investigadores ni profesionales del habla hispana en el área escolar, de tal modo que su difusión se ha limitado a las originales publicaciones en el habla inglesa. Frente al Sistema Koppitz, en el área de la evaluación actual de la habilidad visomotora en niños, uno de los sistemas de calificación más recientes y aún poco conocido en el habla hispana, es el *Sistema de Calificación Cualitativa (SCC)*, Brannigan y

Brunner, 1989; 1996; 2002), que se aplica a la versión modificada del TGB. Esta versión fue modificada para niños entre 4 y 8 años, basándose en las sugerencias de la misma Bender sobre la elección de las láminas y el sistema de puntuación (Brannigan y Brunner, 2002) derivados de la batería de Hirsch (Jansky y deHirsch, 1972). Las investigaciones comparando la eficacia del *SCC* y del Sistema Evolutivo de Koppitz indican que los coeficientes de validez tienden a ser sistemáticamente más elevadas usando el *SCC* (Brannigan y Brunner, 2002). Sin embargo, esto no ha sido replicado en otros grupos de edad o de culturas diferentes.

### *Objetivo*

El objetivo del presente trabajo es evaluar la validez de constructo del *SCC* respecto a una prueba de habilidades pre-académicas para niños del último nivel de preescolaridad y de 1er grado de primaria. La línea de esta investigación es psicométrica respecto a la búsqueda de evidencias de validez, desde métodos multivariados. Establecer la asociación entre una medida de habilidades pre-académicas y el TGB puede aún ser considerado una línea de investigación vigente, pues excepto el análisis de ítems publicado por Merino, (2009), la versión del TGB usada aquí no ha sido anteriormente considerada en alguna investigación iberoamericana; y en segundo lugar, las evidencias de validez con criterios reportadas en los estudios originales (Brannigan y Brunner, 1989, 1996, 2002) se hicieron con instrumentos diferentes a los conocidos en los medios hispanos.

### **Método**

#### *Participantes*

Los niños participantes fueron 70 preescolares (41 niños, 58.6%), entre 5 años y 5 años 5 6 meses de edad, todos provenientes de tres centros educativos preescolares, ubicados en la zona urbana de Lima Metropolitana. La distribución del sexo en los tres centros fue similar a la distribución total en la muestra total. Los colegios se eligieron por el previo contacto entre el investigador y los directores de estos colegios. La disponibilidad de este grupo aseguró la completa participación de los niños y la autorización de los padres de familia durante el periodo de recolección de datos. Son colegios públicos cuya estructura organizativa y funcional pueden ser considerados similares a los demás colegios de gestión gubernamental, respecto al número de alumnos por profesor, distribución de turnos, disponibilidad de servicios básicos educativos, y horas de trabajo pedagógico. Estos centros escolares fueron elegidos intencionalmente.

Las características sociales y culturales de las familias desde los que provienen los niños indican que las madres tienden a dedicar más tiempo en la crianza de los niños, y las labores hogareñas; y en menor proporción, comparten los deberes hogareños con trabajos independientes y eventuales, que generan ingresos por debajo del sueldo mínimo. El nivel educativo de la madre mayormente alcanza la secundaria completa, y es quien centraliza informalmente la preparación escolar y social del durante las horas en que niños no está en el colegio.

### *Instrumento*

Las pruebas seleccionadas como criterios de habilidades pre-académicas corresponden a dos modelos de concepción de las habilidades preparatorias para la lectura: madurativo y conductual (Sélles, 2006). Estas pruebas pueden guardar similitud con las pruebas informales pedagógicas que evalúan habilidades amplias, y cuya estructura y los criterios de calificación son altamente variables. Las capacidades evaluadas por estas pruebas se identifican como habilidades de pre-requisito (Grupo ALBOR-COHS, 2005) o habilidades facilitadoras (Sélles, 2006).

*Test Gestáltico de Bender Modificado (TGB-M, Brannigan y Brunner, 2002).* La versión modificada contiene seis de los nueve diseños originales (A, 1, 2, 4, 6 y 8) para su aplicación a niños preescolares hasta los primeros grados del nivel primario (4.5 hasta 8.5 años), dado que son los más apropiados para niños pequeños. El manual describe un sistema para puntuar el desempeño gráfico del niño, el *Sistema de Calificación Cualitativa, SCC* (Brannigan y Brunner, 2002), de 6 puntos por diseño, desde una puntuación de 0 (líneas aleatorias, garabateo, sin concepto del diseño) hasta 5 (representación exacta del diseño); y que logran gran diferenciación en la evaluación de la calidad los dibujos. Esta versión se califica por un método de inspección global, que refleja el grado de diferenciación y de la gestalt de los diseños reproducidos. La investigación original (Brannigan & Brunner, 2002) y las realizadas independientemente (Fuller y Vance, 1995; Merino, 2009, 2010, en prensa; Merino y Benites, 2011) sobre la confiabilidad interna, test-retest e inter-calificadores y la validez del SCC da soporte a sus propiedades métricas y sus cualidades instrumentales en la evaluación psicopedagógica. Debido a pequeñas diferencias cuantitativas con la forma de aplicación individual del TGB (Caskey y Larson, 1975; Keogh y Smith, 1961), esta versión puede aplicarse también grupalmente. En comparación con el Sistema Evolutivo de Calificación de Koppitz (1984), el SCC muestra correlaciones más elevadas con criterios de rendimiento escolar en diferentes grupos culturales (Brannigan y Brunner, 2002; Chan, 2002). La consistencia interna

( $\alpha$ ) del SCC en el presente estudio fue 0.69.

*Prueba 5 y 6, Forma B* (Gastelumendi, Isasmendi, Slowak y Carbonell, 1969, 1971). Esta es una prueba diseñada para la evaluación de las habilidades preparatorias para el aprendizaje de la lectura, escritura y matemáticas durante el 1er grado de educación regular. Contiene 4 subescalas, de las que se derivan 4 puntajes y también un puntaje general. Las áreas medidas por estas subescalas son *Comprensión* (7 ítems), *Percepción* (12 ítems distribuidos en Parte 1 basado en completar el dibujo de una figura humana; y parte B, ítems de discriminación figura fondo, y orientación espacial), *Precálculo* (5 ítems) y *Motricidad* (5 ítems). Aunque los ítems se califican como 1 o 0, el formato de los ítems es variable respecto al número de alternativas y la forma de presentación de estas, así que puede identificarse como una prueba con cuyas partes son congénicas (Feldt y Brennan, 1989). Se puede considerar a la prueba 5 y 6 como una prueba de estructura y contenido con moderada comunalidad con pruebas pedagógicas para la evaluación rápida de las habilidades en proceso de aprendizaje en clases de preescolar 5 años. Sus adicionales características de facilidad de administración y de calificación lo hace una herramienta que puede ser utilizada por profesores de aula, principalmente. El material estandarizado e impreso de la prueba 5 y 6 es otra de sus ventajas, frente a las pruebas hechas por profesores, pues estas son altamente variables entre instituciones educativas y aun entre profesores de una misma institución. Los autores reportan una relativa mejor especificidad en la detección de problemas de lecto-escritura al inicio de la escolaridad (Isasmendi, A. & Gastelumendi, s/f)

La evaluación de sus propiedades psicométricas para el presente estudio produjo lo siguiente: se eliminaron ítem con bajas correlación ítem-test, con efecto de piso en algunos ítems (ítems excesivamente fáciles) y con varianza cercana a cero. De este modo, se eliminó un ítem (ítem 3; correlación ítem-test corregida [ $r_{itc}$ ] = 0.07) de visomotricidad, dos ítems de Cálculo (ítems 4 y 5;  $r_{itc} < 0.11$ ), tres ítems de *Comprensión* (ítems, 1, 2 y 7;  $r_{itc} < 0.0$ ); un ítem de *Percepción B* ( $r_{itc} < 0.0$ ) y finalmente, tres ítems de la *Figura Humana (FH)* que tuvieron varianza cero. Estas modificaciones mejoraron la consistencia interna y logró mantener un razonable modelo equivalente tau entre los ítems de las subpruebas del 5 y 6, excepto para la subescala *Percepción Visual*. La subescala *Percepción Visual* tiene dos partes: una es el completamiento de una *Figura Humana (FH)*, y la otra lo componen a su vez dos conjuntos de ítems (*Parte A* y *Parte B*) de formatos diferentes: en la *Parte A*, los ítems exploran discriminación Figura-Fondo, la ubicación derecha – izquierda y posiciones espaciales; mientras que en la *Parte B*, los 5 ítems enfatizan la discriminación visual. El puntaje total de esta sub-



escala (Percepción Visual) se calculó promediando ambos componentes *A* y *B* (como lo indica el manual), mientras que el puntaje de la *Figura Humana (FH)* fue la suma simple del número de ítems completados; la correlación entre ambos componentes *A* y *B* fue 0.42 ( $p < 0.01$ ). Ya que esta subescala de *Percepción Visual* se obtuvo sumando el puntaje de *FH* con el promedio de las partes *A* y *B*, esta configuración de dos partes sugiere el uso de un coeficiente ad hoc para esta situación. Ya que la discrepancia entre las medias y varianzas es grande (para *Percepción*,  $M = 3.01$ ,  $DE = 0.88$ ; y para *FH*,  $M = 6.81$ ,  $DE = 2.31$ ), el modelo de medición adecuado para esta subescala es el de partes congénicas (Feldt y Brennan, 1989), por lo tanto la estimación de la consistencia interna más apropiada se obtuvo del coeficiente de Angoff (1953) y Feldt (1975). Esta estimación fue 0.56.

**Tabla 1. Información estadística básica de los puntajes de las utilizadas**

	Min	Max	M	DE	Coef	QSS	5y6	Ca	Co	PV	V
SCC normalizado	12	23	18.32	2.31	0.70 <sup>a</sup>	1					
Puntaje 5 y 6 normalizado	6.75	23.75	16.8272	4.12	0.41 <sup>a</sup> 0.56 <sup>b</sup>	.391 <sup>**</sup>	1				
Calculo (Ca)	.00	3.00	2.1765	.87	0.47 <sup>a</sup>	.175 (0.305 <sup>c</sup> )	.473 <sup>**</sup>	1			
Comprensión (Co)	.00	4.00	2.6912	1.09	0.46 <sup>a</sup>	.067 (0.118 <sup>c</sup> )	.455 <sup>**</sup>	.057	1		
Percepción Visual (PV)	1.50	15.75	9.7978	2.82	0.43 <sup>a</sup> 0.58 <sup>b</sup>	.321 <sup>**</sup> (0.586 <sup>c</sup> )	.864 <sup>**</sup>	.272 <sup>*</sup>	.202	1	
Visomotricidad (V)	.00	4.00	2.1618	1.25	0.52 <sup>a</sup>	.387 <sup>**</sup> (0.641 <sup>c</sup> )	.567 <sup>**</sup>	.217	.178	.256 <sup>*</sup>	1

<sup>a</sup>: Coeficiente alfa. <sup>b</sup>: Coeficiente Angoff-Feldt. <sup>c</sup>: Correlaciones desatenuadas por error de medición.

\*:  $p < 0.05$ , \*\*:  $p < 0.01$

### *Procedimiento y Análisis de datos*

La administración de las pruebas se hizo en condiciones estandarizadas y se mantuvo en lo posible las instrucciones de administración individualizada para asegurar la varianza relevante al constructo durante el proceso de respuesta (Bracken, 2007; Lee, Reynolds y Willson, 2003), incluyendo la introducción de pausas para evitar la fatiga de los niños. Sin embargo, se diseñaron algunas modificaciones en las instrucciones que permitían que los niños entendieran mejor el procedimiento de respuesta y continuación en las tareas presentadas; pero estas modificaciones se aplicaron eventualmente y se consideró que no tuvieron un impacto efectivo y consistente sobre el desempeño de los niños, en el modo que se podría haber dado pistas sobre las respuestas correctas. El 5 y 6, se aplicó de modo individual o grupal; la modalidad individual de administración se hizo con niños cuyo ritmo de trabajo y conducta podrían disminuir el desempeño en la prueba grupal. Por otro lado, la evaluación del TGB se hizo individualmente. Todas las evaluaciones se hicieron en la misma institución escolar de los niños, y en aulas condicionadas para la administración.

Durante la fase de calificación, las tareas que requerían algún tipo de evaluación subjetiva, para decidir la puntuación, se calificaron por un asistente, monitoreando su aplicación de los criterios de calificación de las pruebas; esto sirvió para disminuir la variabilidad por error en la calificación de estas tareas. Esto fue esencialmente cierto para el TGB y para algunas tareas del 5 y 6, como los subtest de Coordinación Visomotora y Figura Humana. En una sesión se entrenó el uso de los criterios de puntuación y se monitoreó su dominio con varios ejemplos.

Los análisis se ejecutaron mediante el programa SYSTAT 12. El procedimiento estadístico considerado fue la correlación canónica entre las medidas. Este procedimiento estima las correlaciones desde procedimientos multivariados, cuando hay un sistema de medidas intercorreladas (Hair, Black, Babin, Anderson y Tatham, 2009). El análisis correlacional usará correlaciones no corregidas y corregidas por atenuación de la confiabilidad de los instrumentos (Nunnally y Bernstein, 1995).

### **Resultados**

Previamente al análisis principal, se examinaron las puntuaciones para detectar posibles irregularidades que limitarían la aplicación del análisis de correlación canónica. Se eliminaron los puntajes de dos sujetos cuyos valores bajos fueron los más extremos, es decir, 3

desviaciones estándares debajo de la media (Sheskin, 2004). Después de la remoción de tales datos, se normalizaron los puntajes del SSC; por otro lado, no fue necesario normalizar los puntajes del 5 y 6 ya que se mantuvieron razonablemente no alejadas de la distribución normal teórica ( $z_{sim}$  y  $z_{cur} < 1.5$ ) y su efecto no sería importante en los cálculos multivariados (Hair et al., 2009).

La variable canónica definida por las subescalas en el 5 y 6 logra explicar el 34.9% de la varianza total; esta estimación en la variabilidad total es debida a la covariación conjunta entre los indicadores con la variable latente del 5 y 6. Sin embargo, esta combinación lineal está claramente más ponderada por las tareas que requieren discriminación visual y habilidad visomotora (ver Tabla 2); y lo es menos con las tareas asumidas para representar la comprensión auditiva y habilidades de pre-cálculo. La correlación corregida por atenuación del error de medición (Nunnally y Bernstein, 1995) entre cada subescala con el puntaje total del 5 y 6 fue para *Cálculo*, *Comprensión*, *Visomotor* y *Precepción Visual* 0.28, 0.22, 0.27 y 0.35, respectivamente.

**Tabla 2. Coeficientes R y canónicos para las subescalas de la prueba 5 y 6**

<i>Prueba 5 y 6</i>	R canónico	Coefficiente Estandarizado
Calculo (Ca)	0.38	0.094
Comprensión (Co)	0.14	-0.089
Percepción Visual (PV)	0.70	0.51**
Visomotricidad (V)	0.85	0.716**

\*\* :  $p < 0.01$

Para las correlaciones atenuadas y no atenuadas, la correlación canónica entre el criterio y el SCC fue 0.45 ( $F_{Rao} [4, 63] = 4.08, p = 0.005$ ) y 0.78 ( $F_{Rao} [4, 63] = 24.66, p < 0.001$ ), respectivamente; en ambos se rechazó la hipótesis nula de asociación aleatoria entre el QSS y los componentes del 5 y 6. Se observa que la magnitud de la correlación canónica con las variables desatenuadas por error de medición es 1.73 veces más grande. Derivado de esta estimación, la magnitud de la correlación multivariada ha sido  $R^2 = 0.20$  ( $R^2_{aj} = 0.15$ ) y  $R^2 = 0.61$  ( $R^2_{aj} = 0.58$ ) para las correlaciones atenuadas y no atenuadas, respectivamente. Como se

observa, el impacto de la desatenuación de las correlaciones ha sido sustancial.

En la Tabla 3 se presenta la correlación canónica entre los componentes del 5 y 6 dentro de su función canónica y los puntajes del SSC, así como sus estadísticos asociados. Hallamos que los componentes del 5 y 6 asociados al desempeño visomotor y percepción visual se relacionan con el SSC en una magnitud relativamente mayor que los otros componentes del 5 y 6, y el efecto del control del error de medición sobre los coeficientes y la prueba *F* fue claramente favorable. La desatenuación por error de medición también reveló que el SSC está asociado a los puntajes en las habilidades de pre-cálculo más allá del error de muestreo, pero en menor magnitud que con los componentes visomotores y perceptuales del 5 y 6.

La falta de asociación multivariada con las tareas de comprensión verbal da soporte a la validez divergente entre esta medida y el SSC. Se debe observar que la covariación entre estas medidas ha sido muy baja y puede considerarse esencialmente cero.

**Tabla 3. Resultados de la correlación canónica atenuada y desatenuada entre las variables canónicas del SCC y prueba 5 y 6**

	R <sup>2</sup> (R <sup>2</sup> <sub>aj</sub> )		Prueba F		Coef. Estandarizado	
	Atenuado	No atenuado	Atenuado	No atenuado	Atenuado	No atenuado
<i>Prueba 5 y 6</i>						
Calculo	0.031 (0.015)	0.093	2.083	6.769*	0.175	0.305*
Comprensión	0.005 (0.0)	0.014	0.300	0.932	0.067	0.118
Percepción Visual	0.103 (0.08)	0.342	7.578**	34.338**	0.32**	0.585**
Visomotricidad	0.150 (0.13)	0.411	11.626**	46.032**	0.387**	0.641**

\*\* : *p* < 0.01

\* : *p* < 0.05

## Discusión

El presente estudio tuvo como finalidad evaluar la relación multivariada entre el *Sistema de Calificación Cualitativa* para la versión modificada del Test Gestáltico de Bender para niños (*TGB-M*, Brannigan y Brunner, 2002) y una medida de rendimiento pre-académico (prueba 5 y 6), en niños de 5 años durante su preparación para los aprendizajes en el primer grado de primaria. El contexto de la investigación fue la evaluación de habilidades pre-académicas de niños, para la identificación temprana de problemas en el futuro aprendizaje escolar. Los resultados mostraron una varianza significativamente mayor entre el *TGB-M* y las tareas de percepción visual y visomotora, y menos varianza compartida del *TGB-M* con las tareas de comprensión y cálculo, tal como fueron medidas por la prueba preparatoria para el primer grado de primaria. Para obtener una más precisa estimación, las correlaciones de desatenuaron por error de medición. Cuando se aplicó este procedimiento de desatenuación, la magnitud de las correlaciones lineales aumentó, y los puntajes de la subescala Cálculo (prueba 5 y 6) demostraron tener una relación sistemática, aunque de baja magnitud. Estos resultados son consistentes con la red nomológica esperada entre el TGB y otras variables de rendimiento estandarizado (Brannigan y Brunner, 2002; Koppitz, 1984; Reynolds, 2007), pues la interrelación de estas habilidades en edades preescolares tiende a ser común. Por ejemplo, Brannigan y Brunner (2002) reportaron correlaciones alrededor de 0.25 con pruebas estandarizadas de rendimiento pre-académico y académico, sugiriendo que la varianza compartida entre el *SCC* y las pruebas psicopedagógicas dan un soporte para incluir el *TGB-M* en la evaluación de preescolares que ingresan al primer grado de primaria.

Por otro lado, hay consideraciones necesarias para comunicar derivadas de los presente resultados. Por ejemplo, la prueba de criterio elegida, el 5 y 6, es una medida estandarizada, breve y de fácil aplicación; pero como criterio unitario del constructo de rendimiento pre-académico, se halló que sus indicadores externos no se integran psicométricamente bien, de tal modo que la variable canónica obtenida recibía la mayoría de su varianza de solo dos sub-test, es decir, Percepción visual y Visomotricidad. En esta misma variable canónica, los ítems de Cálculo contribuyen poco para explicar las variaciones en los puntajes totales de la variable canónica, y menos aún lo hace la subescala Comprensión. Entonces, esta variable canónica captura esencialmente la varianza compartida entre respuestas de los niños a las tareas visuales y visomotoras del 5 y 6. Este contexto psicométrico no hace un favorable respaldo a esta prueba, pues el muestreo del contenido en las tareas de Comprensión y Cálculo no parece

haber contribuído para obtener respuestas consistentes con el atributo latente general que suponen medir. Además de esta limitación en la validez de constructo, el impacto de la varianza de error sobre los puntajes del 5 y 6 ha sido notorio en la magnitud de la prueba de hipótesis nula (prueba  $F$ ) y en los coeficientes estandarizados obtenidos. Ya que la varianza de error limita la magnitud de los coeficientes de validez (McDonald, 1999; Nunnally y Bernstein, 1995), la corrección por atenuación en la relación entre Cálculo y TGB-M produjo que la relación entre ellos aumentará cerca de tres veces la magnitud inicial de la prueba  $F$  y del coeficiente beta estandarizado. Las estimaciones corregidas por atenuación permiten observar las asociaciones entre las variables en una situación ideal de datos libres de error aleatorio, pero finalmente son estimaciones teóricas que el lector debe tomar en cuenta en la interpretación de estos resultados.

Una de las limitaciones de este estudio es que, desde un enfoque estadístico, la manera de elegir a los participantes no asegura la generalizabilidad de los presentes resultados, y únicamente se podrían hacer presunciones sobre su posible extensión hacia otros grupos de participantes que el lector defina como similares. Sin embargo, la recolección de datos y toda la organización de la administración de las pruebas se mimetiza con realidad de la evaluación psicológica y educativa. Este contexto fue la exploración de habilidades pre-académicas en grupos de menos de 30 niños, dentro de una institución escolar que requiere evaluaciones para obtener descripciones del rendimiento o identificación temprana de problemas de rendimiento académico. Consideramos que la generalizabilidad de los presentes resultados, como un aspecto de la validez, se puede evaluar como una cuestión de grado y no de una afirmación categórica, y esta posición es congruente con lo expresado por los *Estándares* (AERA et al., 1999) respecto a las evidencias de validez.

Aunque la meta de la presente investigación fue aportar la covariación multivariada entre un sistema de calificación para el TGB-M y un criterio de rendimiento preescolar pre-académico, consideramos que es necesario abrir camino hacia la comparación entre otros sistemas para calificar el TGB. La validez comparativa frente a otros sistemas de calificación es aún menos frecuente que el reporte de la covariación de una sistema de calificación (Lacks, 1999), y el presente reporte contribuye a disminuir esta limitación. Sin embargo, la contribución de los presentes resultados puede servir como una línea base respecto a un sistema de calificación que exige una evaluación global de los diseños reproducidos, y que ofrece un

nuevo enfoque de calificación de una prueba psicológica tan tradicional y popular como el TGB.

Frente a otros sistemas de calificados diseñados para una evaluación discreta de los diseños, los sistemas cualitativos han tendido a demostrar más varianza compartida con criterios relevantes (Brannigan y Decker, 2003; Chan, 2000; Parsons y Weinberg, 1993), pero reconocemos que la utilidad de un nuevo dispositivo de evaluación se expresa mejor en un contexto comparativo frente a otras medidas existentes (Haynes, Nelson y Blaine, 1999; Hunsley y Meyer, 2003), de tal modo que la valoración de la validez tendría un referente empírico que sería más útil al futuro usuario.

Las limitaciones expuestas no son criterios que podrían disminuir la importancia psicométrica de los presentes resultados, pues con los parámetros reportados aquí pueden garantizar la disponibilidad de datos empíricos para una replicación, estableciendo la magnitud de la validez hallada como una hipótesis nula contrastable. El sistema de Koppitz (1984), de Sugar (Parsons y Weinberg, 1993), de Puntuación Global (Sisto, Noronha, y Santos, 2005), y los más recientes aportes de Brannigan y Decker (2003) y Reynolds (2007), pueden ser candidatos para establecer la validez comparativa del sistema de calificación elegido en este artículo. Estas versiones no han sido evaluadas en otros medios culturales diferentes al de su construcción, y su efectividad sólo se ha presentado en publicaciones anglosajonas.

Se debe anotar que los resultados obtenidos podrían poner en cuestionamiento el aporte que podría dar el TGB-M cuando se lo usa junto con una batería de pruebas que ya tuviera entre sus componentes otra medida de habilidad visomotora. En efecto, de acuerdo a los resultados hallados, en que se encontró una varianza compartida significativa y exclusiva entre el TGB-M y tal componente visomotor, se podría pensar que incluir el TGB-M sería redundante en una batería que ya contiene una subescala visomotora, y que tal subescala puede ser suficiente para propósitos de despistaje. Pero es más probable hallar baterías de pruebas cuyas subescalas se diseñen para hacer un muestreo de contenido muy incompleto comparándolo con una prueba ad hoc para evaluar la integración visomotora, como el TGB-M. Además de criterios psicométricos para seleccionar una prueba, hay criterios pragmáticos también importantes, como el costo monetario, la curva de su aprendizaje para su administración y calificación, y su portabilidad y su validez incremental (Haynes et al., 1999; Hunsley y Meyer, 2003); por lo tanto, las baterías pueden ser menos exhaustivas en la medición de alguna habi-



alidad debido a su naturaleza multidimensional. Y dado que la habilidad visomotora contribuye a explicar un monto de varianza aceptable del aprendizaje en los primeros ciclos de educación escolar (Beery, 2000; Kulp, 1999; Kulp et al., 2004), entonces la inclusión de una versión moderna y psicométricamente satisfactoria del TGB puede ser de valor práctico para la evaluación psicopedagógica. En una de estas aplicaciones del TGB-M podría servir de línea base para contrastar los cambios implementados por actividades rutinarias, planificadas o iniciadas por el niño dentro un programa de intervención psicomotriz (Candel, 2005).

## Referencias

- AERA, APA & NCME. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Angoff, W. F. (1953). Test reliability and effective test length. *Psychometrika*, 18, 1-14.
- Beery, K. E. (2000). *Prueba Beery-Buktenica del desarrollo de la integración visomotriz* (4 ed.). México, D.F.: El Manual Moderno.
- Berdicewski, O., & Milicic, N. (2004). *Prueba de Funciones Básicas (35ta. ed.)*. Santiago: Editorial Universitaria.
- Bracken, B. (2007). Creating the optimal preschool testing situation. In B. Bracken & Nagle (Eds.) *Psychoeducational assessment of preschool children* (pp. 137-153). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Brannigan, G. G., & Brunner, N. A. (1989). *The Modified Version of the Bender-Gestalt Test for Preschool and Primary School Children*. Brandon, VT: Clinical Psychology.
- Brannigan, G. G., & Brunner, N. A. (1996). *The Modified Version of Bender-Gestalt Test for Preschool and Primary School Children-Revised*. Brandon, VT: Clinical Psychology Publishing.
- Brannigan, G. G., & Brunner, N. A. (2002). *Guide to the Qualitative Scoring System for the modified version of the Bender-Gestalt Test (2nd ed.)*. IL: Charles C. Thomas.
- Brannigan, G. G., & Decker, S. L. (2003). *Bender Visual-Motor Gestalt Test (2nd ed.)*. Itasca, IL: Riverside Publishing.
- Brannigan, G. G., & Decker, S. L. (2006). The Bender-Gestalt II. *American Journal of Orthopsychiatry*, 76, 10-12.
- Brannigan, G. G., Decker, S. L., & Madsen, D. H. (2004). *Innovative features of the Bender-Gestalt II and expanded guidelines for the use of the global scoring system. (Bender Visual-Motor Gestalt Test, 2nd ed., Assessment Service Bulletin No. 1)*. Itasca, IL: Riverside Publishing.
- Candel, I. (2005). Elaboración de un programa de atención temprana. *Electronic Journal of Research of Educational Psychology*, 7(3), 151-192.
- Caskey, W. R., Jr. & Larson, G. L. (1975). Two modes of administration of the Bender Visual-Motor Gestalt Test to kindergarten children. *Perceptual and Motor Skills*, 45(1), 1003-1006.
- Chan, P. W. (2000). Comparison of visual motor development in Hong Kong and USA as-

- sessed on the Qualitative Scoring System for the Modified Bender Gestalt Test. *Psychology Reports*, 88, 236-240.
- Chan, P. W. (2002). Relationship of the visual motor development and academic performance in young children in Hong Kong assessed in the Bender-Gestalt Test. *Perceptual and Motor Skills*, 90, 209-214.
- Cummings, J. A., Hoida, J. A., Macheck, G. R., & Nelson, J. M. (2003). Visual-motor assessment of children. En C. R. Reynolds & R. W. Kamphaus (Eds.) *Handbook of psychological and educational assessment of children: intelligence, aptitude, and achievement* (2nd. Ed., pp. 498-518). New York: Guilford Press.
- DeHirsch, K., Jansky, J.J., & Langford, W.S. (1966). *Predicting reading failure*. New York: Harper and Row.
- Feldt, L. S. (1975). Estimation of reliability of a test divided into two parts of unequal length. *Psychometrika*, 40, 557-561.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.
- Filho, L. (1960). *Los Tests ABC* (6ta ed.). Buenos Aires: Kapeluz.
- Fuller, G. B. & Vance, B. (1995). Interscorer reliability of the modified version of the bender-gestalt test for preschool and primary school children. *Psychology in the Schools*, 32(4), 264-266.
- Gastelumendi, E., & Isasmendi, A. (s/f) Factores que favorecen la lecto-escritura visto a través del Test 5-6. Documento no publicado: Autores.
- Gastelumendi, E., Isasmendi, A., Slowak, G., & Carbonell, Z. (1968). *Test 5-6, Forma A: Exploración de las áreas fundamentales al iniciar el aprendizaje de las materias básicas*. Documento no publicado: Autores.
- Gastelumendi, E., Isasmendi, A., Slowak, G., & Carbonell, Z. (1969, Marzo). *El Test 5-6*. Ponencia presentada en el XII Congreso Interamericano de Psicología, Montevideo, Uruguay.
- Gastelumendi, E., Isasmendi, A., Slowak, G., & Carbonell, Z. (1971). El Test 5-6, Forma B: Exploración de las áreas fundamentales al iniciar el aprendizaje de las materias básicas. *Enfoques de Aula*, 1, 20-38.
- Grupo Albor-Cohs (2005). *Batería Magallanes de habilidades Básicas de Aprendizaje (BAMADI-vm)*. Madrid: Grupo ALBOR-COHS.

- Hair, J., Black, B., Babin, B., Anderson, R., & Tatham, T. (2009). *Multivariate Data Analysis* (7<sup>th</sup> edition). Upper Saddle River, NJ; Prentice-Hall.
- Haynes, S. N., Nelson, K., & Blaine, D. C. (1999). Psychometric issues in assessment research. In P. C. Kendall, J. N. Butcher & G. N. Hombeck (Eds.) *Handbook in research methods in clinical psychology* (pp. 125-154). New York: Wiley.
- Hunsley, J., & Meyer, G. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological and statistical issues. *Psychological Assessment*, 15(4), 446-455.
- International Test Commission (ITC) (2000). *Guidelines on Test Use: Spanish Version*. Translation authorised by the Colegio Oficial de Psicólogos de España. ITC: Author.
- Jansky, J. & deHirsch, K. (1972). *Preventing reading failure: Prediction, diagnosis, intervention*. New York: Harper and Row.
- Johnston, C. W., Lanak, B. (1985). Comparison of the Koppitz and Watkins Scoring Systems for the Bender Gestalt Test. *Journal of Learning Disabilities*, 18(7) 377-378.
- Keogh, B. K., & Smith, C. E. (1961). Group techniques and proposed scoring system for the Bender-Gestalt Test with children. *Journal of Clinical Psychology*, 17, 172-175.
- Koppitz, E. M. (1984). *El Test Guestáltico Visomotor para niños* (10ma. Ed.). Buenos Aires: Guadalupe.
- Kulp, M. T. (1999). Relationship between visual motor integration skill and academic performance in kindergarten through third grade. *Optometry and Vision Science*, 76, 159-163.
- Kulp, M. T., Earley, M. J., Mitchell, G. L., Timmerman, L. M., Frasco, C. S., & Geiger, M. E. (2004). Are visual perceptual skills related to mathematics ability in second through sixth grade children? *FOCUS on Learning Problems in Mathematics*, 26(4), 44-51.
- Lacks, P. (1999). *Bender Gestalt screening for brain dysfunction* (2nd. ed.) New York: John Wiley.
- Lee, D., Reynolds, C. R., & Willson, V. L. (2003). Standardized test administration: Why bother? *Journal of Forensic Neuropsychology*, 3, 55-81.
- McDonald, R. P. (1999). *Theory test: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Merino, C. & Benites, L. (2011). Evaluación de la confiabilidad en dos grupos de edad, usando el Sistema Cualitativo de Calificación para el Test de Bender Modificado. *Universitas Psicológica*, 10, 237-249.
- Merino, C. (2009). Un análisis no paramétrico de ítems de la Prueba Gestáltica de Bender

- Modificada para estudiantes de primaria. *Liberabit*, 15(2), 23-43.
- Merino, C. (2010). El Sistema de Calificación Cualitativa para la Prueba Gestáltica de Bender – Modificada: Estudio preliminar de sus propiedades psicométricas. *Avances en Psicología Latinoamericana*, 28(1), 63-73.
- Merino, C. (En prensa). Validez comparativa de tres sistemas de calificación para el Test Gestáltico Visomotor de Bender. *Revista de Psicología – Universidad César Vallejo*.
- Nunnally, J. C., & Bernstein, I. J. (1995). *Teoría psicométrica* (3ra. ed.). México, D. F.: McGraw-Hill.
- Palisano, R. L. & Dichter, C. G. (1989). Comparison of two tests of visual-motor development used to assess children with learning disabilities. *Perceptual and Motor Skills*, 68(3), 1009-1103.
- Parsons, L., & Weinberg, S. L. (1993). The Sugar Scoring System for the Bender-Gestalt. *Perceptual and Motor Skills*, 77, 883-893.
- Perticone, E. X. (1998). *The Clinical and Projective Use of the Bender–Gestalt Test*. Springfield, IL: Thomas
- Preda, C. (1997). Test of visual-motor integration: Construct validity in a comparison with the Beery-Buktenica Developmental Test of Visual-Motor Integration. *Perceptual and Motor Skills*, 84(3/2), 1439-1443.
- Reichenberg, N. & Raphael, A. J. (1992). *Advanced psychodiagnostic interpretation of the Bender Gestalt test: Adults and children*. New York: Praeger.
- Reynolds, C. R. (2007). *Koppitz Developmental Scoring System for the Bender Gestalt Test: Examiner's manual (2nd ed.)*. Austin, TX: Pro-Ed.
- Sélles, P. (2006). Estado actual de la evaluación de los predictores y de las habilidades relacionadas con el desarrollo inicial de la lectura. *Aula Abierta*, 88, 53-72.
- Sheskin, D. (2004). *Handbook of parametric and nonparametric statistics* (3<sup>rd</sup> ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Simner, M. L. (1991). Estimating a child's learning potential from form errors in a child's printing. In J. Wann, A.M. Wing, & N. Sovik (Eds.), *Development of graphics skills: Research, perspectives, and educational implications* (pp. 205-222). London: Academic Press Inc.
- Simner, M. L., & Barnes, M. J. (1991). Relationship between first-grade marks and the high school dropout problem. *Journal of School Psychology*, 29, 331-335.
- Sisto, F. F., Noronha, A. P. P., Santos, A. A. A. (2005). *Teste Gestáltico Visomotor de Bender: Sistema de Pontuação Gradual (B-SPG)*. São Paulo: Vetor

- Telegdy, G. A. (1974b). The relationship between socioeconomic status and school readiness. *Psychology in the Schools, 11*, 351-356.
- Telegdy, G. A. (1975). The effectiveness of four reading tests as predictors of first grade achievement. *Psychology in the Schools, 12*, 4-11.
- Telegdy, G. A. (1974a). A factor analysis of four school readiness tests. *Psychology in the Schools, 11*(2), 127-133.
- Wallbrown, J. D., Engin, A. W., Wallbrown, F. H., & Blaha, J. (1975). The prediction of first grade reading achievement with selected perceptual-cognitive tests. *Psychology in the Schools, 12*(2), 140-149.
- Wallbrown, J. D., Wallbrown, F. H. & Engin, A. W. (1974). The relative importance of mental age and selected assessors of auditory and visual perception in the metropolitan readiness test. *Psychology in the Schools, 11*(2), 136-143.
- Watkins, E. O. (1976). *The Watkins Bender Gestalt Scoring System*. Novato, CA: Academic Therapy.