

## VALIDEZ COMPARATIVA DE TRES SISTEMAS DE CALIFICACIÓN DEL TEST GESTÁLTICO VISOMOTOR DE BENDER

César Merino Soto\*

Universidad de San Martín de Porres

### RESUMEN

Se presenta un estudio preliminar que compara coeficientes de validez entre tres versiones de calificación del Test Gestáltico Visomotor de Bender (TGB) y puntajes de rendimiento preacadémico. Estas versiones el sistema de Calificación Cualitativa (SCC), el sistema de Puntuación Gradual (SPG) y el Sistema Evolutivo de Koppitz (SEK). La muestra fueron niños que ingresaban al primer grado de primaria. Las puntuaciones del TGB fueron ajustadas a un modelo no paramétrico en la Teoría de Respuesta al Ítem, y las correlaciones fueron parcializadas por la edad y la habilidad intelectual. Se obtuvo que los sistemas de enfoque ordinal tuvieron correlaciones más elevadas; especialmente el SSC, que más con las demás, y que además retuvo más varianza con la edad. Se discute la interpretación de estos resultados, así como las limitaciones y la importancia práctica.

*Palabras claves:* Bender Gestalt Test, validez, visomotor, evaluación.

### COMPARATIVE VALIDITY OF THREE SYSTEMS FOR THE BENDER GESTALT TEST

#### ABSTRACT

In this study, it is presented a preliminary study of comparison of validity coefficients between three scoring systems for the Bender Gestalt Test (TGB) and pre-academic achievement and intellectual ability scores. These versions are: Qualitative Rating System (SCC), the Gradual Scoring system (SPG) and Koppitz's Developmental Scoring System (SEK). The sample were composed by children who enter to the first grade. TGB scores were adjusted to a nonparametric model in Item Response Theory, the correlations were controlled by age and intellectual ability. It was found that ordinal systems approach had higher correlations, and also retained more variance with age. It was discussed the results interpretations the limitations and practical importance.

*Key words:* Bender Gestalt Test, validity, visual-motor, assessment

### VALIDADE COMPARATIVO DE TRÊS SISTEMAS DE CLASSIFICAÇÃO DE BENDER GESTALT TESTE

#### RESUMO

Nós apresentamos um estudo preliminar comparando os coeficientes de validade entre três versões do Teste Bender Gestalt pontuação Bender (TGB) e dezenas de pré-acadêmica realização. Estas versões Sistema de Avaliação Qualitativa (SCC), Sistema de Avaliação Gradual (GPS) e Sistema de Koppitz Evolucionária (SEK). Os shows foram crianças que ingressam na primeira série. TGB pontuações foram ajustados a um modelo não paramétrico na Teoria de Resposta ao Item, e as correlações foram influenciados pela idade e capacidade intelectual. Verificou-se que os sistemas de abordagem ordinal tiveram correlações mais altas, especialmente a SSC, a maioria com os outros, e também reteve mais variância com a idade. Discutimos a interpretação desses resultados e as limitações e importância prática.

*Palavras-chave:* Teste Bender Gestalt, validade, visual-motor, avaliação.

\*Docente adscrito al Instituto de Investigación de Psicología de la Universidad San Martín de Porres y docente en la Universidad Científica del Sur. Correspondencia: sikayax@yahoo.com.ar

El Test Gestáltico Visomotor de Bender (TGB, Bender, 1938) no necesita mucha presentación entre los psicólogos investigadores y aplicados, pues por décadas está dentro de los instrumentos más utilizados (Archer & Newsom, 2000; Sullivan & Bowden, 1997). Las variaciones de esta prueba ponen retos a los usuarios respecto al entrenamiento e interpretación de sus resultados (Dana, Field & Bolton, 1983). Una de estas variaciones estructurales fue la versión abreviada de solo seis diseños, aplicada a niños pequeños y derivada del trabajo de deHirsch, Jansky y Langford (1966). La versión abreviada 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 estas versiones (Telegdy, 1974a; Telegdy, 1974b; Telegdy, 1975; Wallbrown, Engin, Wallbrown & Blaha, 1975; Wallbrown, Wallbrown & Engin, 1974). Los estudios anteriores sobre la forma modificada sugieren su efectiva contribución para predecir el estatus de rendimiento académico, así como los puntajes de pruebas estandarizadas de rendimiento en un tiempo posterior. Estos estudios se han efectuado en población norteamericana y sus hallazgos podrían ser limitados a las características de medición y de la muestra evaluada.

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, por ejemplo, el habla inglesa. Frente al *Sistema Evolutivo de Koppitz (SEK)*, (1984), en el área de la evaluación actual de la habilidad visomotora en niños hay un sistema de calificación más reciente y aún poco conocido en el habla hispana: el *Sistema de Calificación Cualitativa (SCC)*, (Brannigan & 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 & Brunner, 2002) derivados de los trabajos de deHirsch (deHirsch et al., 1966; Jansky & deHirsch, 1972) y de Keogh y Smith (1961). Las investigaciones comparando la eficacia del *SCC* y del Sistema Evolutivo de calificación de Koppitz indican que los coeficientes de validez tienden a ser sistemáticamente más elevadas usando el *SCC* (Brannigan & Brunner, 1991; 2002; Chan, 2002). De estos estudios se puede identificar que el enfoque ordinal del *SCC* captura mejor los cambios evolutivos en el desempeño medido por el TGB.

En Latinoamérica, específicamente en Brasil, ha aparecido recientemente un sistema de calificación cuyo proceso de elaboración se fundamentó en la adaptación de la formulación original de Koppitz, enfatizando el grado de distorsión de los diseños reproducidos. Este sistema, el Sistema de Puntuación Gradual (*SPG*, Sisto, Noronha & Santos, 2006), usó métodos basados en la teoría clásica de los test y la teoría de respuesta al ítem. Propuso un sistema ordinal que pudiera expresar mejor las variaciones en el desempeño visomotor y ha demostrado tener mejores propiedades psicométricas que el *SEK* (Noronha & Mattos, 2006).

Las versiones de calificación del TGB se incluyen en la evaluación psicológica como un soporte para explorar el estado actual del niño; sin embargo, su efectividad debería ser puesta a prueba mediante la comparación entre varias modalidades de calificación del TGB. Por este motivo, el objetivo de la presente investigación es comparar la validez de los puntajes de tres versiones de calificación de TGB, respecto a un criterio de habilidad intelectual y de rendimiento pre-académico. Estos sistemas de calificación se derivan de dos conceptualizaciones diferentes del modo de calificación del desempeño visomotor: uno de naturaleza continua y evaluación global de los diseños (*SCC*), y los otros de naturaleza discretas y basados en la puntuación del error (*SEK* y *SPG*). Esta comparación es exploratoria y en pequeña muestra, y permitiría observar el comportamiento psicométrico de estos sistemas

de calificación en las condiciones habituales en que puede ocurrir la evaluación psicológica en el ejercicio profesional; es decir, en pequeñas muestras y con alternativos métodos que en combinación pueden ayudar a tomar decisiones sobre las características de desarrollo del niño evaluado.

## MÉTODO

### Participantes

Los participantes de la muestra fueron 32 niños y niñas, de una edad promedio de 62.4 meses (DE = 3.97 meses; Mín. = 62 meses; Máx. = 80 meses), todos provenientes de varias instituciones educativas preescolares de gestión estatal, en la etapa de ingreso al primer grado de educación primaria (entre octubre y febrero del siguiente año escolar). Las características básicas de este grupo aparecen en la Tabla 1. El distrito de procedencia de la muestra tiene principalmente áreas urbanizadas y es considerado como una población menos pobre (quintil 4), y sus indicadores de pobreza son menos severos que otros distritos en la provincia Lima (FONCODES, 2006). Dado que las características funcionales, estructurales y organizacionales de estas instituciones educativas preescolares son similares entre ellas, se puede asumir una similitud en el monto y ritmo de experiencias de aprendizaje de los niños participantes frente a los de la población en Lima. Debido a la zona de ubicación de los colegios, las familias de los niños alcanzan mayormente el nivel de instrucción secundario básico, las madres tienden a pasar más horas con el niño, se ocupan del hogar y eventualmente realizan actividades económicas independientes; y mayoritariamente, las familias de los niños conviven con otros miembros de la familia extendida. Por lo tanto, los hogares de los niños lo integran generalmente más de tres miembros, con padres de condición civil de casados o convivientes, y pertenecientes a la clase media o menos, percibiendo alrededor de 1000 nuevos soles mensuales, o menos.

La habilidad cognitiva general de los niños

se distribuyó levemente con asimétrica negativa, según los resultados de una prueba de inteligencia (ver la sección Instrumentos).

### Instrumentos

*Sistema de Calificación Cualitativa para el Test Gestáltico de Bender Modificado* (Brannigan & Brunner, 1989, 1996, 2002).

Esta es la versión del TGB que usa únicamente seis de los diseños originales (A, 1, 2, 4, 6 y 8) para su aplicación en niños preescolares hasta los primeros grados del nivel primario (4.5 hasta 8.5 años), dado que tales diseños son los más apropiados para niños pequeños. Incluye un sistema para puntuar el desempeño gráfico del niño, el *Sistema de Calificación Cualitativa* (SCC, Brannigan & Brunner, 2002) de 6 puntos, desde una puntuación de 0 (líneas aleatorias, garabateo, sin concepto de los diseños) hasta 5 (representación exacta del diseño) y que logran una gran diferenciación en la evaluación de los dibujos. Esta versión se califica por un método de inspección global, que refleja el grado de diferenciación y la gestalt de los diseños reproducidos. La investigación sobre la confiabilidad interna, test-retest e inter-jueces, y la validez del SCC generalmente da soporte a su valor métrico y su utilidad en la evaluación psicopedagógica (Brannigan & Brunner, 2002). Frente al Sistema Evolutivo de Calificación de Koppitz, el SCC muestra correlaciones más elevadas con criterios de rendimiento escolar (por ejemplo, Brannigan & Brunner, 2002; Chan, 2002). El manual presenta una extensa revisión de los hallazgos psicométricos, así como los criterios de calificación de cada diseño. La consistencia interna (coeficiente alfa) de los puntajes del SCC en la muestra fue 0.88.

*Prueba de Dibujo de la Figura Humana para Estimación Intelectual en Niños, Adolescentes y Adultos* (DAP: IQ, Reynolds & Hickman, 2004).

Esta es la más reciente prueba basada en el dibujo de una persona para estimar la habilidad intelectual aplicable en personas de 4 años hasta adultos (89 años). La instrucción para administrarlo consiste en solicitarle al examinado

Tabla 1  
*Información demográfica de los participantes*

	N	%
Sexo		
Varón	19	59.4
Mujer	13	40.6
Orden de nacimiento		
1ro	15	46.9
2do	11	34.4
3ro	4	12.5
4to	2	6.2
Ocupación de la madre		
Casa	24	75.0
Trabajo	8	25.0
Instrucción de los padres		
Madre (Padre)		
Primaria incompleta	2 (1)	6.2 (3.1)
Primaria completa	2 (1)	6.2 (3.1)
Secundaria incompleta	9 (5)	28.1 (15.6)
Secundaria completa	10 (13)	31.2 (40.6)
Técnico incompleta	4 (4)	12.5 (12.5)
Técnico completa	3 (6)	9.4 (18.8)
Universidad incompleta	1 (1)	3.1 (3.1)
Universidad completa	1 (1)	3.1 (3.1)
Estado civil (Madre)		
Casada	10	31.2
Conviviente	20	62.5
Separada	1	3.1
Divorciada	1	3.1
Convivencia familiar		
Ambos padres	29	
Solo un padre	2	
Con otros	1	

que se dibuje a sí mismo; las instrucciones estandarizadas de calificación del *DAP:IQ* enfatizan los aspectos conceptuales del dibujo y no la calidad artística. El material consiste en un lápiz, hoja Bond A-4 y la hoja de registro. El sistema de calificación se basa en una puntuación cualitativa que varía desde 0 (parte no reconocible, incompleta o ausente) hasta 4 puntos

(parte presente o cumpliendo todos los criterios anteriores); esta variación depende del ítem calificado. Los ítems forman 23 criterios estandarizados para calificar el dibujo: siete criterios son calificados con 0 y 1; siete criterios con 0, 1 y 2; ocho criterios con niveles de 0, 1, 2 y 3; finalmente, un criterio con niveles de puntuación desde el 0 hasta el 4. El puntaje se



obtiene de la suma simple de las puntuaciones alcanzadas en los criterios; luego se deriva un puntaje de CI, que es una estimación cuantitativa y confiable de la habilidad cognitiva general. El manual se reporta que los puntajes del *DAP: IQ* producen resultados estables ( $r > 0.80$ ) en cortos periodos de tiempo (una semana) en una muestra de 45 personas de 6 a 57 años. Un reciente estudio en el ambiente latinoamericano (Dávila, Tello & Merino, 2009) han confirmado que la consistencia interna es replicable ( $\geq 0.79$ ), así como el acuerdo intercalificadores e intracalificadores usando correlaciones intraclase en escolares ( $\geq 0.80$ ) y preescolares, ( $\geq 0.72$ ). Los estudios reportados en el manual mencionan que las correlaciones de validez concurrente con el Sistema Koppitz y Goodenough-Harris fueron 0.85 y 0.86, respectivamente; y con el WISC-III,  $r_{prom} = 0.51$ ; y en el estudio latinoamericano, las correlaciones concurrentes con el Sistema Koppitz fueron elevadas,  $r_{prom} = 0.81$  (Dávila et al., 2009); en este mismo estudio, las correlaciones con una prueba de razonamiento no verbal fueron más elevadas que con el Koppitz. La muestra normativa original del *DAP: IQ* fueron 3090 personas de las variadas regiones de los Estados Unidos. Este instrumento cumplirá la función de ayudar a establecer la validez divergente (validez de constructo). En los presentes puntajes, la consistencia interna estimada con el coeficiente Gilmer-Feldt fue 0.82.

*Sistema Evolutivo de Koppitz (SEK, Koppitz, 1984).*

Es un sistema basado en la calificación dicotómica de 30 errores en la reproducción de los diseños, y clasificados en errores de distorsión de la forma, rotación, sustitución, perseveración y desintegración. Se aplica entre los 5 y 11 años de edad. Es uno de los más populares para la evaluación de la madurez visomotora en niños (Brito, 2001; Lacks, 1999). Koppitz (1984) reportó numerosos estudios sobre su capacidad discriminativa entre niños normales frente a estudiantes con problemas de conducta, de aprendizaje, problemas médicos y lesionados cerebrales, así como su validez de constructo con

pruebas de inteligencia, rendimiento escolar y habilidades preacadémicas. Desde sus inicios, uno de sus usos ha sido orientado hacia el despistaje problemas de rendimiento escolar y de aprendizaje (Keogh, 1969). Las estimaciones de confiabilidad test-retest ( $> 0.70$ ) e intercalificadores ( $> 80\%$  de acuerdo) reportadas por Koppitz (1984) fueron satisfactorias. Por otro lado, hay una extensa producción científica publicada sobre la validez de esta versión del TGB, se hallaron resultados contradictorios, especialmente evidencias de la no aplicabilidad de las normas originales de Koppitz en otros contextos socioculturales (Billingslea, 1963; Cummings, Hoida, Machek & Nelson, 2003; Keogh, 1969). Las primeras normas para este sistema se obtuvieron con 1104 niños (Koppitz 1963, 1984). El coeficiente de consistencia interna en los puntajes de la presente muestra fue 0.65.

*Sistema de Puntuación Gradual (SPG, Sisto, Noronha & Santos, 2006).*

El *SPG* evalúa la maduración perceptual-motora en niños entre 6 y 10 años. Fue creado en Brasil luego de que Sisto et al (2006) revisaran las numerosas investigaciones efectuadas en su país con el *SEK*. Se halló una inconsistente relación entre el *SEK* y criterios de rendimiento escolar en Brasil, así como problemas de unidimensionalidad (Bartholomeu & Sisto, 2008). Fundamentalmente, el *SPG* califica el grado de distorsión de la forma en los diseños reproducidos mediante puntajes que varían entre 0 y 3. Estudios previos han informado sobre la convergencia entre el *SEK* y el *SPG*, con correlaciones lineales mayores a 0.80 (Noronha & Mattos, 2006); por otro lado, la confiabilidad por consistencia interna (entre 0.69 y 0.75) y el acuerdo entre calificadores (entre 85% y 87%) reportados en el manual indican valores apropiados (Sisto et al., 2006). El *SPG* comparte más varianza compartida con el rendimiento académico y la edad que con el *SEK* (Noronha & Mattos, 2006). En la presente muestra, la consistencia interna (alfa de Cronbach) fue 0.82.

Tabla 2  
*Información descriptiva de los puntajes de las pruebas aplicadas.*

	Min	Máx	M	DE
Batería de Despistaje Pre-Académico General				
Reconocimiento de Letras y Palabras	1	18	8.25	5.67
Habilidades Fonológicas	0	17	8.69	5.00
Percepción Visual	0	21	12.81	6.34
Habilidades Cuantitativas	0	23	14.31	7.16
Vocabulario/Conceptualización	0	20	13.47	5.01
BDPG – Total	1	94	57.53	25.95
Sistemas de calificación para el TGB				
Sistema de Calificación Cualitativa (SCC)	14.60	24.55	19.42	3.13
Sistema Evolutivo de Koppitz (SEK)	3.29	14.55	7.96	3.04
Sistema de Puntuación Gradual (SPG)	3.43	12.93	7.93	3.12
Dibujo de una Persona (DAP: IQ)	4	27	16.28	5.79

*Batería de Despistaje para Primer Grado* (Merino, 2010).

Este instrumento explora habilidades preacadémicas para niños que ingresan al primer grado de primaria; las habilidades evaluadas están en un estrecho rango de contenido sobre Conocimiento de Letras y Palabras (18 ítems), Habilidades Fonológicas (17 ítems), Percepción Visual (21 ítems), Habilidades Cuantitativas (23 ítems) y Habilidades de Vocabulario y Conceptualización (20 ítems). Para cada una de estas áreas se obtiene un puntaje derivado de la suma simple de los ítems correctamente respondidos, además de un puntaje total basado en la suma de las subescalas. Los ítems son de tipo opción múltiple, entre tres y cuatro opciones de respuesta. Exceptuando la tarea de Seguimiento de Instrucciones, todas las demás usan estímulos gráficos. Las evidencias de confiabilidad muestran coeficientes KR-20 corregidos por

Horst mayores a 0.79; también, la validez de la estructura interna y su relación con criterios de rendimiento académico ha sido satisfactoria (Merino, 2010).

### **Procedimiento**

Considerando que se ha reportado su equivalencia con la modalidad de administración individual (Koppitz, 1984; Buckley 1978; Tolor y Brannigan, 1980), la versión abreviada del TGB-M se administró también grupalmente, entre 6 a 12 niños por grupo, y en orden balanceado respecto a los otros instrumentos. Se mantuvieron constantes las condiciones de administración grupal estandarizada para minimizar la varianza de error proveniente del proceso de recolección de datos (Bracken, 2000; McCallin, 2006) y en concordancia con las directrices para el uso apropiado de pruebas (Hambleton, 1996;

International Test Commission, 2000).

La calificación de las pruebas las efectuó el autor y una asistente estudiante de psicología, quien recibió un prolongado periodo de entrenamiento en el uso de los sistemas de calificación para el BGT. La mitad de los protocolos fueron revisados para verificar el acuerdo de los puntajes, resultando alrededor de un 95% de acuerdo entre ambos calificadores en los tres sistemas de calificación. El análisis correlacional usó la puntuación promedio de los ítems obtenida de ambos calificadores, para así lograr una estimación más confiable del puntaje.

El análisis correlacional se efectuará por correlaciones Pearson de orden cero y correlaciones lineales parciales de 2do orden. Las variables de control serán la edad (medido en meses) y el puntaje de habilidad intelectual como es medido por el DAP: IQ. Considerando el tamaño de la muestra y la inestabilidad de los resultados consecuente con esta limitación, se aplicarán unos métodos que proponen dar más información a las estimaciones de habilidad de los evaluados, derivados de la suma de los puntajes a los ítems. Aun cuando la suma de las respuestas a los ítems es el más directo indicador del nivel de atributo del evaluado (Ramsay, 1995), el presente estudio usará estimaciones más precisas de la habilidad evaluada. Estas estimaciones se obtendrán desde un modelo no paramétrico de teoría de respuesta al ítem (TRI), propuesto por Ramsay (1991), y que aplica una técnica basada en el análisis funcional de datos (Ramsay, 1991; Ramsay & Silverman, 2005) que son estimaciones no paramétricas directas de las funciones de respuesta al ítem y que contrastan con los métodos TRI usados por el modelado paramétrico. Los modelos TRI no paramétricos emplean toda la información disponible de las respuestas a los ítems para hacer una estimación del atributo latente que subyace al puntaje directo, prescindiendo de un modelo a priori (Sachs, Law, Chan & Rao, 2001). Los cálculos necesarios para las estimaciones a nivel de los ítems y del puntaje en el atributo latente se hicieron por medio del programa *Testgraf* (Ramsay, 1995; 2000). La aplicación apropiada de este modelo requiere la

determinación de un ajuste para calcular la función de respuesta del examinado, sustentado en las técnicas *kernel*. Estos métodos usan un parámetro ( $h$ ) de suavizamiento (*smooth*) de curvas distribucionales para un mejor ajuste de los datos a una función de distribución no paramétrica. El parámetro de suavizamiento  $h$  sugerido por *Testgraf* fue 0.55 (calculado como  $h = 1.1 N^{-\frac{1}{5}}$ ).

Las estimaciones de habilidad se aplicarán para todos los puntajes usados. Por otro lado, ya que los sistemas *SCC* y *SPG* usan un procedimiento graduado de puntuación, las estimaciones de los parámetros de los ítems tienden a más precisos y podrían demandar menos ítems para los puntajes de atributo latente  $\theta$  (Ramsay, 2000); en la situación del *SEK*, con ítems dicotómicos, esta regla no se aplica. Pero la cantidad de sus indicadores (18 ítems o indicadores) permite aceptar los parámetros estimados por *Testgraf*, ya que se sugiere que un instrumento contenga 15 o más ítems (Ramsay, 2000). En el análisis correlacional del presente estudio, se utilizarán estimaciones del atributo latente (habilidad visomotora) en la misma métrica del puntaje total que varió entre 2.5 desviaciones estándares alrededor de la media.

## RESULTADOS

### *Convergencia entre los sistemas de calificación.*

Primeramente, se observan las asociaciones lineales entre los sistemas de calificación. Las correlaciones convergentes ajustadas por edad y habilidad intelectual fueron bajas respecto a sus magnitudes no ajustadas, pero superiores a |0.50| (Tabla 3). Adicionalmente, se puede detectar un patrón en las magnitudes de estos coeficientes. El puntaje de *SCC* correlaciona más elevado con *SPG* y menos con *SEK*, y este último correlaciona algo más elevado con *SPG* que con el *SCC*. Este patrón puede sugerir la comunalidad de algunos aspectos específicos que se puntúan en los sistemas, como la exactitud de la copia versus puntuación de errores o el grado de cuantificación del puntaje ordinal versus puntaje discreto.

Tabla 3  
Correlaciones entre los sistemas de calificación y criterios: *r* parcial y correlación de primer orden (entre paréntesis)

	SCC	SEK	SPG
Sistemas de calificación			
Sistema de Calificación cualitativa (SCC)	1		
Sistema Evolutivo de Koppitz (SEK)	-0.536 (-0.694)	1	
Sistema de Puntuación Gradual (SPG)	-0.639 <sup>a</sup> (-0.781)	0.591 <sup>a</sup> (0.776)	1
Variables control			
Edad en meses	0.522 <sup>a</sup>	-0.231	-0.338
DAP:IQ	0.552 <sup>a</sup>	-0.664 <sup>a</sup>	-0.673 <sup>a</sup>
Variables criterio			
Reconocimiento de Letras y Palabras	0.542 <sup>a</sup> (0.722)	-0.508 <sup>a</sup> (-0.731)	-0.482 <sup>a</sup> (-0.729)
Habilidades Fonológicas	0.373 <sup>b</sup> (0.609)	-0.446 <sup>b</sup> (-0.713)	-0.354 (-0.67)
Percepción Visual	0.354 (0.608)	-0.239 (-0.551)	-0.126 (-0.507)
Habilidades Cuantitativas	0.477 <sup>c</sup> (0.679)	-0.378 <sup>b</sup> (-0.644)	-0.417 <sup>b</sup> (-0.68)
Vocabulario/Conceptualización	0.345 (0.594)	-0.301 (-0.527)	-0.325 (-0.57)
BDPG – Total	0.520 <sup>a</sup> (0.72)	-0.44 <sup>b</sup> (-0.71)	-0.41 <sup>b</sup> (-0.71)

<sup>a</sup>:  $p < 0.008$  (p ajustado con el método Bonferroni)

<sup>b</sup>:  $p < 0.05$

<sup>c</sup>:  $p = 0.008$

#### Correlaciones con criterios externos.

Las correlaciones no corregidas por el efecto de la edad y la inteligencia no verbal fueron más elevadas, como era de esperarse (Tabla 3), ya que una parte de la varianza total correspondiente a estas covariables aún no fue removida para estimar mejor las correlaciones. Luego de la remoción de estas covariables, nuestros resultados indican que el SSC tiende a compartir más varianza común con el criterio de habilidades preacadémicas, en lugar que los sistemas basados en errores. Exceptuando el puntaje de las habilidades de discriminación fonológica, la varianza compartida con los puntajes de SSC ha sido alrededor de 24%, mientras que para el SEK y SPG fue 19% y 16% de la varianza compartida, respectivamente.

El patrón de las correlaciones parciales entre los sistemas basados en errores (SEK y SPG) y los

puntajes criteriosales ha sido bastante similar y ligeramente mayor para el SEK; sin embargo, ambos han tendido a ser menores respecto al SCC.

El monto de diferencia entre las correlaciones de primer orden y las correlaciones parcializadas mostró que los sistemas basados en puntuaciones graduadas retuvieron más varianza aún después de controlar los efectos de la edad y la habilidad intelectual. Esto fue particularmente observable en el SSC, pues la varianza mediana aún retenida en este sistema fue mayor (6.19%) frente al SEK (5.4%). la varianza retenida en el SPG fue 5.21%, una valor entre las dos magnitudes anteriores.

## DISCUSIÓN

En las pocas investigaciones hispanas sobre las habilidades visomotoras, una de las áreas



menos frecuentemente estudiadas en la validez de las pruebas visomotoras es la comparación de sus correlaciones con constructos (Lacks, 1999); por lo tanto, el objetivo del presente estudio fue estimar y comparar varios sistemas de calificación para una versión abreviada del TGB, específicamente el Sistema de Calificación Cualitativa (*SCC*, Brannigan y Brunner, 2002), el Sistema de Puntuación Gradual (*SPG*, Sisto et al., 2006) y el Sistema Evolutivo de Koppitz (*SEK*, Koppitz, 1984). Observando las correlaciones concurrentes entre los sistemas de calificación, los resultados convergen con el patrón de las estimaciones correlacionales en muestras americanas (Brannigan & Brunner, 1991; Parsons & Weinberg, 1993; Schachter, Brannigan & Tooke, 1991) y en una muestra china (Chan, 2002); es decir, las correlaciones entre los sistemas en las edades de la muestra de estudio son elevadas y mayormente sobre 0.50. Sin embargo, en los estudios citados no se controlaron los efectos de la edad ni de la habilidad intelectual. En la muestra de estudio, luego de parcializar los efectos de la edad y de la habilidad cognitiva no verbal, las correlaciones bajaron sustancialmente, pero aun así las asociaciones explicaban un significativo monto de varianza compartida en niveles considerados, montos que se pueden considerar de nivel moderado o elevado (Cohen, 1992). Otro patrón existente de correlaciones fue que los dos sistemas con una puntuación ordinal (*SCC* y *SPG*) muestran correlaciones algo más elevadas, y que el sistema ordinal que califica el similitud de las reproducciones con respecto a los diseños (*SCC*), mostró correlaciones de mayor magnitud que los sistemas que puntúan los errores en el desempeño visomotor (*SPG* y *SEK*). Este patrón es consistente con los aspectos conceptuales y técnicos de estos sistemas, pero sugiere que el grado de asociación está influenciado por el modo de escalamiento y por el aspecto que se califica en el desempeño visomotor. La diferencia de la varianza separada de los puntajes debido a la edad y habilidad intelectual con las medidas de visomotricidad fueron parecidas, pero algo menores en el *SCC*. Esto parece sugerir que el puntaje del *SCC* incorpora un poco más al efecto de la edad en la

variabilidad de sus puntajes, y por lo tanto podría ser más sensible a la edad que el *SEK* y *SPG*. Sin embargo, se requiere una confirmación de esto en una muestra más grande para obtener más estabilidad en los resultados.

Las magnitudes de estas correlaciones están en la línea de los resultados entre medidas de visomotricidad, en que la varianza compartida generalmente entre ellas supera alrededor 40% (por ejemplo, Beery, 2000; Brannigan & Brunner, 2002; Brannigan & Decker, 2003; Sisto et al., 2006). Pero también se reconoce cierta varianza no explicada que indica que los sistemas de calificación proporcionan información diferente, correspondiendo a aspectos tales como varianza específica y varianza de error. Estas variables van más allá de la influencia de la edad y la habilidad intelectual, y podrían referirse a las habilidades atencionales, la habilidad motora o la capacidad perceptual visual.

Otro aspecto revelado es que el *SCC* mostró que las correlaciones entre sus puntajes y las variables de criterio estudiadas (habilidades preacadémicas) reportan más varianza común, frente a dos sistemas enfocados en un diferente marco de evaluación del desempeño visomotor. Esto indica que un sistema de calificación orientado hacia la “bondad de ajuste” entre el diseño y la respuesta de copiado logra recuperar más información y permite más variabilidad lineal que los sistemas en que se observa el error, la distorsión del diseño o la “maldad del ajuste” entre el diseño y la copia realizada.

Estudios previos han cuestionado sistemas de calificación que enfatizan el error de las reproducciones y en la observación atomística de las reproducciones en el test de Bender (Brannigan & Brunner, 2002; Brannigan & Decker, 2003; Decker, 2008b; Keogh & Smith, 1961; Moose & Brannigan, 1997) u orientada a “signos” (Keogh, 1969). En convergencia con ello, los resultados permiten considerar que el marco de puntuación construido para valorar el grado de exactitud entre el estímulo y su copia tiene un mejor poder discriminativo con criterios de rendimiento preacadémico, además de capturar

mejor los cambios producidos por desarrollo de la habilidad visomotora que los sistemas objetivos (Brannigan & Brunner, 1989, 1991, 1996, 2002; Decker, 2008a, 2008b; Moose & Brannigan, 1997).

Los resultados presentados deben ser atenuados por un aspecto que impone sus limitaciones, es decir el tamaño muestral, pues esto compromete la generalización y estabilidad de los estadísticos obtenidos, y se requeriría más tamaño muestral para detectar significancia estadística en varias de las correlaciones halladas (Cohen, 2007). Aunque varias de las correlaciones halladas no fueron estadísticamente significativas, la magnitud de ellas generalmente no ha sido pequeña, sino moderada; la significancia estadística es fuertemente afectada por el tamaño muestral (Cohen, 2007). En contrapartida con esta limitación, también debe reconocerse que la magnitud de las correlaciones ha estado entre las halladas frecuentemente en la literatura correlacional entre los puntajes del BGT y criterios de rendimiento académico; así, la similitud de nuestros resultados con los reportados puede indicar evidencias de una aceptable replicación.

Finalmente, aunque la incorporación de estimaciones no paramétricas en el marco de la teoría de respuesta al ítem puede ofrecer estimaciones más precisas de la habilidad de los sujetos, aún se considera necesario la extensión de este estudio en una muestra más grande para obtener mayor poder estadístico y protegerse contra error Tipo II (Cohen 2007). Pero la limitación respecto al tamaño muestral no remueve una importancia práctica de este estudio: que calificaciones ordinales pueden ser teóricamente mejor implementadas e interpretadas que los sistemas basados en error, y demostrar mejor validez de constructo en la evaluación de los cambios evolutivos (Decker, 2008a; 2008b) debido a que no subestiman las correlaciones. De ello puede deducirse que si el profesional o investigador requiere crear un sistema de calificación, podría tener una mejor justificación implementar un sistema ordinal y aproximado de la naturaleza continua del

constructo medido.

El estudio preliminar presentado puede ser una referencia relevante en la validez de constructo de diferentes versiones del TGB no comparadas anteriormente y que pueden ser consideradas como instrumentos modernos en la práctica evaluativa, considerando que cada versión pone retos en el entrenamiento, la interpretación y validación (Dana et al., 1983) y un énfasis diferente en la evaluación de la visomotricidad. Otra justificación para usar versiones modernas del TGB es respecto a las normas de interpretación. La información disponible sobre las normas publicadas en medios científicos latinoamericanos indica que muchas de ellas tienen un existencia de más de 15 años (Brito, Alfradique, Pereira, Porto & Santos, 1998; Brito & Santos, 1996; Casullo, 1988; Chang, 1990; Kroeff, 1992, 1988), y como actualmente hay versiones del TGB con una conceptualización distinta al modelo de Koppitz que incluyen métodos psicométricos modernos, debería de cuestionarse el uso del Sistema evolutivo de Koppitz como una referencia fundamental para la práctica normativa del desarrollo visomotor actual, aunque con continuo valor clínico.

## REFERENCIAS

- Archer, R. P. & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment*, 7(3), 227-235.
- Beery, K. E. (2000). *Prueba Beery-Buktenica del Desarrollo de la Integración Visomotriz*. México, DF: El Manual Moderno.
- Bartholomeu, D. & Sisto, F. F. (2008). Maturidade viso-motora e inteligência: Um estudio correlacional. *Psicologia: Ciência e profissão*, 28(2), 362-373.
- Bender, L. (1938). A visual-motor gestalt test and its clinical use. Research Monographs, No. 3, New York: American Orthopsychiatric Association.
- Billingslea, F. Y. (1963). The Bender Gestalt: A review and perspective. *Psychological Bulletin*, 60, 233-251.

- Bracken, B. A. (2000). Maximizing content-relevant variance: The assessment situation. En B. A. Bracken (Ed.) *Psychoeducational assessment of preschool children, 3th. ed.* (pp. 33-44). Needham Heights, MA: Allyn & Bacon.
- Brannigan, G. A. & Brunner, N. A. (1991). Relation between two scoring systems for the modified version of the Bender Gestalt Test. *Perceptual and Motor Skills, 72*, 286.
- Brannigan, G. G. & Brunner, N. A. (2002). *Guide to the Qualitative Scoring System for the modified version of the Bender-Gestalt Test*. Springfield, IL: Charles C. Thomas.
- 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 Publishing.
- Brannigan, G. G., & Brunner, N. A. (1996). *The Modified Version of the Bender-Gestalt Test for Preschool and Primary School Children-Revised*. Brandon, VT: Clinical Psychology Publishing.
- Brannigan, G. G., & Decker, S. L. (2003). *Bender Visual-Motor Gestalt Test, Second Edition*. Itasca, IL: Riverside Publishing.
- Brito, G. N. (2001). Bender Gestalt. In W. E. Craighead, & C. B. Nemeroff (Eds.) *The Corsini Encyclopedia of Psychology and Behavioral Science, 3th edition* (Vol. 1, pp. 201-203). New York: Wiley and Sons.
- Brito, G. N. & Santos, T. R. (1996). The Bender Gestalt Test for 5- to 15-year old Brazilian children: Norms and validity. *Brazilian Journal of Biological Research, 29*(11), 1513-1518.
- Brito, G. N., Alfradique, G. M., Pereira, C. C., Porto, C. M. & Santos, T. R. (1998). Developmental norms for eight instrument used in the neuropsychological assessment of children: Studies in Brazil. *Brazilian Journal of Biological Research, 31*(3), 399-412.
- Buckley, P. D. (1978). The Bender Gestalt Test: A review of reported research with school-age subjects, 1966-1977. *Psychology in the Schools, 15*(3), 327-338.
- Casullo, M. M. (1988). *Test de Bender Infantil. Normas Regionales Argentinas*. Buenos Aires.: Guadalupe.
- 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.
- Chang, G. (1990). *Nueva Escala de Maduración del Bender Infantil*. Lima: Biblioteca Andina de Psicología.
- Cohen, B.H. (2007). *Explaining psychological statistics* (3<sup>a</sup> ed.). New York: John Wiley & Sons.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.
- Cummings, J. A., Hoida, J. A., Machek, 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* (2<sup>a</sup> ed., pp. 498-518). New York: Guilford Press.
- Dana, R. H., Field, K., & Bolton, B. (1983). Variations of the Bender-Gestalt Test: Implications for training and practice. *Journal of Personality Assessment, 47*(1), 76-84.
- Dávila, M., Tello, K., & Merino, C. (2009). Nueva versión del Dibujo de una Persona para la Estimación Intelectual: Estudio psicométrico preliminar en preescolares. *Fractal: Revista de Psicología, 21*(2), 443-444.
- Decker, S. L. (2008a). Measuring growth and decline in visual-motor process with the Bender-Gestalt Second Edition. *Journal of Psychoeducational Assessment, 26*(1), 3-15.
- Decker, S. L. (2008b). Intervention psychometrics: Using norm-referenced methods for treatment planning and monitoring. *Assessment for Effective Intervention, 34*(1), 52-61.
- deHirsch, K., Jansky, J. J., & Langford, W. S. (1966). *Predicting reading failure*. New York: Harper and Row
- FONCODES (2006). *Focalización geográfica: Nuevo mapa de pobreza de FONCODES 2006*. Lima: Ministerio de la Mujer y Desarrollo Social.
- Hambleton, R. K. (1996). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes

- de error, posibles soluciones y directrices practicas. En J. Muñiz (Coord.), *Psicometría* (pp. 207-238). Madrid: Universitas.
- International Test Commission (ITC) (2000). *Guidelines on Test Use: Spanish Version*. Translation authorised by the Colegio Oficial de Psicólogos. ITC: Author.
- Jansky, J., & deHirsch, K. (1972). *Preventing reading failure*. New York: Harper Row.
- Keogh, B. K. (1969). The Bender Gestalt with children: Research implications. *Journal of Special Education*, 3, 15-22.
- 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. (1963). *The Bender Gestalt Test for young children*. New York: Grunne & Strraton.
- Koppitz, E. M. (1984). *El Test Guestáltico Visomotor para niños (10ma. Ed.)*. Buenos Aires: Guadalupe.
- Kroeff, P. (1988). Normas brasileiras para o Teste de Bender. *Psicologia: Reflexão e Crítica*, 1/2 (3), 12-19.
- Kroeff, P. (1992). Desempenho de crianças no Teste de Bender e nível sócioeconômico-cultural. *Psicologia: Reflexão e Crítica*, 5(2), 119-126.
- Lacks, P. (1999). *Bender Gestalt screening for brain dysfunction (2nd ed.)*. New York: John Wiley & Sons.
- McCallin, R. C. (2006). Test Administration. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 625-652). Mahwah, NJ: Lawrence Erlbaum.
- Merino, C. (2010). *BDPG: Bateria de Despistaje para Primer Grado*. Documento no publicado. Lima: Autor.
- Moose, D., & Brannigan, G. G. (1997). Comparison of preschool children's scores on the modified version of the Bender-Gestalt Test and the Developmental Test of Visual-Motor Integration. *Perceptual and Motor Skills*, 85(2), 766-775
- Noronha, A. P. P. & Mattos, R. (2006). Koppitz e Bender - sistema de pontuação gradual: Comparação entre sistemas de avaliação. *Psicologia Escolar e Educational*, 10(2):223-233.
- Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer-Verlag.
- Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56(4), 611-630.
- Ramsay, J. O. (1995). *TESTGRAF: A program for the graphical analysis of multiple choice test and questionnaire data*. Montreal: McGill University.
- Ramsay, J. O. (2000). *TESTGRAF: A program for the graphical analysis of multiple choice test and questionnaire data*. Montreal: McGill University.
- Reynolds, C. R., & Hickman, J. A. (2004). *Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults (DAP: IQ)*. Austin, TX: Pro-Ed.
- Sachs, J., Law, Y. K., Chan, C. K., & Rao, N. (2001). A nonparametric item analysis of the Motivated Strategies for Learning Questionnaire-Chinese Version. *Psychologia – An International Journal of Psychology in the Orient*, 44(3), 197-208.
- Schachter, S., Brannigan, G. G., & Tooke, W. (1991). Comparison of two scorings systems for the modified version of the Bender-Gestalt test. *Journal of School Psychology*, 29(3), 265-269.
- Sisto, F. F., Noronha, A. P. P., & Santos, A. A. A. (2006). *Teste Gestaltico Visomotor de Bender – Sistema de Pontuação Gradual (B-SPG)*. Sao Paulo, SP: Vetor.
- Sullivan, K. & Bowdem, S. C. (1997) Which tests do neuropsychologist use? *Journal of Clinical Psychology*, 53, 657-661.
- Telegdy, G. A. (1974b). The relationship between socioeconomic status and school readiness. *Psychology in the Schools*, 11, 351-356.
- Telegdy, G. A. (1974a) A factor analysis of four school readiness tests. *Psychology in the Schools*, 11(2), 127-133.
- Telegdy, G. A. (1975). The effectiveness of four reading tests as predictors of first grade achievement. *Psychology in the Schools*, 12,

- 4-11.
- Tolor, A. & Brannigan, G. G. (1980). *Research and clinical applications for the Bender-Gestalt Test*. Springfield, IL: Charles C. Thomas Publishers.
- 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.

**Recibido:** 8 de diciembre del 2010

**Aceptado:** 18 de diciembre del 2010