

Dimensionalidad del Test Gestáltico Visomotor de Bender Modificado: un análisis factorial confirmatorio

César Merino Soto¹

Instituto de Investigación de Psicología, de la Universidad de San Martín de Porres, Lima, Perú

RESUMEN

El propósito de este estudio es examinar las características estructurales de una nueva y breve versión del Test Gestáltico Visomotor de Bender, modificada para niños entre 4 y 8 años (TGB-M). Los participantes fueron 1369 niños provenientes de varias zonas urbanas de Lima, Perú. Se aplicó el TGB-M en forma grupal e individual, y se verificó la dimensionalidad mediante la metodología del análisis factorial confirmatorio, y la consistencia interna con metodología factorial y con la teoría clásica de los testes. Los resultados justifican la unidimensionalidad del puntaje total, pero la varianza retenida y la confiabilidad del constructo varía en cada nivel de edad; también, la invarianza configuracional de los ítems es satisfactoria en todos los grupos de edad. La consistencia interna es satisfactoria. Se discute la interpretación de la variabilidad de los resultados entre las edades, así como la posible extensión del TGB-M en niños de más de 8 años.
Palabras clave: sistema de calificación cualitativa; test gestáltico de Bender; análisis factorial; validez; confiabilidad.

ABSTRACT – Dimensionality of Bender Gestalt Visual Motor Test Modified: a confirmatory factor analysis

The aim of this study is to examine the structural characteristics of a new brief version of the Bender Gestalt Test of Bender, modified for children between 4 and 8 years (BGT-M). The participants were 1369 children from various urban areas of Lima, Peru. The BGT-M was group or individually administered, and dimensionality was verified using the methodology of confirmatory factor analysis, and the internal consistency with factorial methodology with the classical test theory methodologies. The results justify the unidimensionality of the total score, but the variance retained and construct reliability varies by age level; also, the configurationally invariance of the items is satisfactory in all age groups. Internal consistency is satisfactory. We discuss the interpretation of the variability of results between ages, and the possible extension of the TGB-M in children over 8 years.

Keywords: qualitative scoring system; Bender gestalt test; factor analysis; validity; reliability.

RESUMO – Dimensionalidade do Teste Gestáltico Visomotor de Bender Modificado: uma análise fatorial confirmatória

O objetivo deste estudo é examinar as características estruturais de uma nova e breve versão do Teste Gestáltico Visomotor de Bender, modificada para crianças entre 4 y 8 anos (TGB-M). Participaram 1369 crianças provenientes de várias zonas urbanas de Lima, Peru. Foi aplicado o TGB-M em grupos e individualmente, e foi verificada a dimensionalidade mediante metodologia de análise fatorial confirmatória, e a consistência interna com metodologia fatorial e com a teoria clássica dos testes. Os resultados justificam a unidimensionalidade da pontuação total, mas a variância retida e a confiabilidade do construto variam em cada nível de idade; também, a invariância configuracional dos itens é satisfatória em todos os grupos de idade. A consistência interna é satisfatória. É discutida a interpretação da variabilidade dos resultados entre as idades, assim como a possível extensão do TGB-M em crianças com mais de 8 anos.
Palavras-chave: sistema de pontuação qualitativa; Teste Gestáltico de Bender; análise fatorial; validade; confiabilidade.

El Test Gestáltico Visomotor de Bender (TGB; Bender, 1938) posiblemente es aún una de las herramientas de evaluación más conocida entre los psicólogos (Decker & Carboni, 2011; Walrath, 2011). Fue originalmente construido para evaluar el funcionamiento visomotor como una expresión de diversas patologías relacionadas con el aprendizaje, integridad neurológica y personalidad. Sobre ese instrumento se han hecho numerosas investigaciones de varios aspectos de su validez, pero una observación de las revisiones de sus estudios

(Billingslea, 1963; Buckley, 1978; Lacks, 1999) llegaría a la conclusión que las investigaciones menos frecuentes han sido sobre la dimensionalidad latente de su puntaje. Efectivamente, los estudios anglosajones sobre la dimensionalidad del TGB han sido escasos comparados con los de la predictividad de sus puntajes y su convergencia con otras medidas visomotoras; y hasta la fecha poco se conoce sobre las características estructurales en revistas publicadas (un trabajo no publicado aparece, sin embargo, en Merino & Allen, 2012), en cualquiera de las versiones

¹ Endereço para correspondência: Instituto de Investigación de la Universidad de San Martín de Porres, Dirección de correspondencia, Av. Tomás Marsano, 242 (5to piso), Lima 34, Perú. E-mail: sikayax@yahoo.com.ar

existentes. En estudios anglosajones, los reportes de la dimensionalidad del TGB consistieron en explorar las relaciones entre los ítems en población clínica (Guertin, 1952, 1954a, 1954b) o en población normal (Brannigan & Decker, 2003; Haynes, 1970); mientras que otros estudios exploraban la comunalidad de los puntajes del TGB con otras medidas para revelar su red nomológica como evidencia de validez de constructo (Arbit & Zager, 1978; Decker, Allen & Choca, 2006; Haynes, 1970).

Los usos del TGB han sido variados: como medida de personalidad en adultos y niños (por ejemplo, respectivamente: Hutt, 1975; Koppitz, 1963, 1975), y como un medio de detección de patologías orgánicas en el cerebro (Lacks, 1999; Groth-Marnat, 2003); pero la interpretación más aceptada de sus puntajes es como una medida de habilidad visomotora, la misma que tiene suficientes evidencias para poder realizar una válida descripción de esta habilidad y para incrementar la validez de las evaluaciones (Brannigan & Brunner, 2002; Decker e cols., 2006). La medición de la habilidad visomotora puede lograrse con un puntaje que represente la variabilidad del desempeño del evaluado, pero se deben obtener evidencias para respaldar la interpretación unidimensional de tales puntajes, y el análisis factorial es una respuesta estadística a esta necesidad.

En una de sus recientes innovaciones, la 2da versión del TGB (Bender-II, Brannigan & Decker, 2003) reporta un análisis factorial exploratorio de los ítems en la muestra de estandarización, dividida en dos grupos de edad (menos de 8 años y 8 años o más). En los grupos las cargas factoriales estuvieron entre 0,61 y 0,84, todas consideradas de elevada magnitud. La varianza retenida por los grupos fue 64,7, 47,5, 50,3 y 54,9, dando una evidencia de la unidimensionalidad del Bender-II. Las diferencias en el monto de varianza retenida fueron relacionadas con la variabilidad de la dificultad de los ítems en cada grupo de edad.

Los resultados factoriales de algunas de otras versiones del TGB para adultos han mostrado inconsistentes conclusiones sobre las características estructurales; por ejemplo, se ha señalado que una dimensión no muestra suficiente varianza para justificar el uso de un solo puntaje interpretativo (por ejemplo, Guertin, 1952, 1954a, 1954b); mientras otros estudios demuestran que la unidimensionalidad de los puntajes de BGT es recomendable para interpretar el desempeño visomotor (Haynes, 1970; Sadeghi & Hooman, 1999). En niños, los resultados de Brannigan y Decker (2003) para el Bender-II indican también una satisfactoria confirmación de la unidimensionalidad. La contradicción de esos resultados puede deberse a los diferentes sistemas de calificación, que enfatizan diferentes aspectos del desempeño y lo cuantifican con sistemas de calificación poco comparables entre sí (Merino, 2011), o por aspectos metodológicos problemáticos, como el uso de correlaciones Pearson en lugar de correlaciones tetracóricas entre los ítems, número de

ítems en relación al número de sujetos, factores de dificultad del ítem y determinación del número de factores para retener.

Uno de los sistemas de calificación más recientes y apenas conocido en el habla hispana, es el Sistema de Calificación Cualitativa (SCC; Brannigan & Brunner, 2002), que se aplica a la versión modificada del TGB para niños entre 4 y 8 años. Esa modificación se fundamentó en las sugerencias de Bender sobre la elección de las láminas (se removieron 6 láminas consideradas difíciles para los niños) y de los trabajos derivados de deHirsch (Jansky & deHirsch, 1972) en niños de los primeros años para la detección temprana de problemas en el aprendizaje. Las investigaciones comparativas entre el SCC y el Sistema Evolutivo de Koppitz (1963, 1975) indican que las evidencias de validez tienden a ser más satisfactorias con el SCC (Brannigan & Brunner, 2002; Chan, 2002; Merino, 2011; Parsons & Weinberg, 1993), pues las correlaciones con los criterios de aprendizaje son mayores comparadas con el método de Koppitz. Sin embargo, la validez de la estructura interna no fue evaluada por algún estudio con el SCC.

La investigación sobre la dimensionalidad de un puntaje provee el soporte empírico de la estructura interna intentada para un instrumento, y permite justificar la interpretación de un puntaje mediante una o varias posibles dimensiones. En ese marco, el objetivo del presente estudio es examinar la dimensionalidad del TGB-M, mediante un enfoque SEM (*Structural Equations Modeling*) aplicado a los puntajes obtenidos del Sistema de Calificación Cualitativa (Brannigan & Brunner, 2002). En los anteriores estudios en población hispana y anglosajona con ese sistema, no se ha efectuado alguna investigación psicométrica al respecto. Es probable que, dado las características de brevedad de los ítems (6 láminas en esta versión modificada del TGB), el igual escalamiento, y la homogeneidad de las tareas (copiado de diseños geométricos), no podrían haber fuertes argumentos para pensar que el puntaje obtenido del TGB-M/SCC pueda representar más de una prominente dimensión latente. Sin embargo, la unidimensionalidad es una hipótesis falsificable cuya comprobación es fundamental para aplicar válidamente otros procedimientos cuantitativos y para interpretar significativamente los puntajes (Messick, 1995).

Método

Participantes

Los participantes fueron 1381 niños(as) de diferentes zonas de la provincia de Lima Metropolitana, en Perú. La selección de la muestra fue no probabilística, obteniéndose acceso a ella mediante la disponibilidad de participación y la aceptación de cartas de invitación y de descripción de un proyecto de adaptación del TGB-M a varias instituciones educativas de primaria dentro de

Lima Metropolitana. En el periodo de evaluación, ninguno de los colegios elegidos participaba de programas de inclusión de personas con discapacidades; tampoco, de actividades con apoyo institucional externo o programas especiales de estimulación de la motricidad o percepción visual.

La descripción socioeconómica de los participantes se hizo mediante la identificación de quintiles de pobreza de cada distrito al que pertenecía la institución educativa muestreada, obtenidos de un documento de trabajo del Fondo de Cooperación para el Desarrollo Social para establecer el mapa de pobreza en el Perú; los quintiles de pobreza son categorías de probabilidades, y van del quintil 1 (distritos con muchas carencias) al 5 (distritos con muy pocas carencias). En la Tabla 1 aparece la descripción demográfica de los participantes. La distribución del sexo para la muestra total fue similar (Tabla 1), así como respecto a la edad ($\chi^2[11] = 15,28, p > 0,05$) y grado escolar ($\chi^2[4] = 5,35, p > 0,05$).

Tabla 1
Distribución demográfica de los participantes

Edad	N	%
4,0m-4,5m	50	3,6
4,6m-4,11m	58	4,2
5,0m-5,5m	96	7,0
5,6m-5,11m	217	15,7
6,0m-6,5m	168	12,2
6,6m-6,11m	110	8,0
7,0m-7,5m	126	9,1
7,6m-7,11m	120	8,7
8,0m-8,5m	149	10,8
8,6m-8,11m	133	9,6
9,0m-9,5m	106	7,7
9,6m-9,11m	48	3,5
Sexo		
Varón	738	53,4
Mujer	643	46,6
Grado escolar		
Inicial 4	93	6,7
Inicial 5	116	8,4
1er grado	564	40,8
2do grado	317	23,0
3er grado	291	21,1
Tipo de colegio		
Público	1018	73,7
Privado	76	5,5
Parroquial	287	20,8
Quintiles de pobreza		
Quintil 3	60 (0,68)	4,4
Quintil 4	825 (0,69)	60,3
Quintil 5	484 (0,71)	35,4

Instrumento

Test Gestáltico de Bender Modificado (TGB-M; Brannigan & 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). Se califica con un sistema para puntuar el desempeño gráfico del niño, el Sistema de Calificación Cualitativa (SCC; Brannigan & Brunner, 2002), de 6 puntos por diseño, desde 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. Esa 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 sobre la confiabilidad y la validez del SCC da soporte a sus propiedades métricas para la evaluación psicopedagógica (Allen, 1992; Brannigan & Brunner, 2002; Fuller & Vance, 1995). El TGB-M se puede administrar grupalmente, ya que se hallan solo diferencias pequeñas con la modalidad de administración individual (Brannigan & Brannigan, 1995; Caskey & Larson, 1975). Frente al Sistema Evolutivo de Calificación de Koppitz (1963, 1975), el SCC ha demostrado validez de predictividad relevante con criterios de rendimiento escolar (Brannigan & Brunner, 2002; Chan, 2002), y los puntajes muestran satisfactorios niveles de confiabilidad por consistencia interna y acuerdo inter-examinadores. Las pocas investigaciones hispanas con el TGB-M/SCC han demostrado también satisfactorias propiedades psicométricas (Merino, 2009, 2010, 2011; Merino & Benites, 2011).

Procedimiento

Una vez obtenida la aprobación del Comité de Ética de la universidad del autor, y completadas las autorizaciones correspondientes, la administración de las pruebas se hizo en condiciones estandarizadas y manteniendo en lo posible las instrucciones de administración grupal o individual para maximizar la varianza relacionada con el constructo medido (Bracken, 2007; Lee, Reynolds, & Willson, 2003). Los evaluadores recibieron entrenamiento para la aplicación grupal e individual; y aunque la modalidad de administración preferida fue grupal, la elección del modo de aplicación se determinó por las condiciones y facilidades estructurales de evaluación encontradas en cada centro educativo que permitió el acceso.

En la aplicación grupal se utilizó un cuadernillo de 6 páginas, en que cada página contenía un diseño impreso en el tercio superior, manteniendo el tamaño original de las tarjetas individuales. Cada grupo de 3 a 5 asistentes evaluó cerca de 20 a 25 niños en cada grupo, en sus respectivas aulas de clase. La actividad se presentó como una tarea de dibujo, requiriendo el mejor desempeño posible en el copiado. En los grupos que se realizó la aplicación individual, las instrucciones fueron las mismas.

Respecto a la calificación, se ha sugerido que el aprendizaje y aplicación del SSC es relativamente sencillo (Allen, 1992), pero se consideró necesario planificar el entrenamiento y supervisión del uso de SCC. Se realizaron entre dos a tres sesiones de entrenamiento para cada grupo de asistentes, en las que se explicó el enfoque global de calificación, y se calificaron con supervisión entre promedio de 10 a 15 protocolos.

Para los análisis, además de los estadísticos descriptivos para los puntajes en las edades de la muestra, se calcularon coeficientes de confiabilidad α (Cronbach, 1951) asumiendo un modelo equivalente tau. Debido al efecto de la restricción del rango de los puntajes sobre el cálculo del coeficiente α , se aplicó la corrección de Gulliksen (1950)² para atenuar este efecto en cada rango de edad.

Para el modelamiento de los datos, primero se examinó la presunción de unidimensionalidad latente de los ítems, mediante la extracción de la varianza retenida por el primer autovalor (Hair, Anderson, Tatham, & Black, 1988); también, se aplicó el gráfico *scree test* (Cattell, 1966) y el análisis paralelo (Horn, 1965). Para ambos procedimientos se usó el programa *Explorer* (Lorenzo-Seva & Ferrando, 2006).

En una segunda fase del análisis, se usó el enfoque *Structural Equations Modeling* (SEM), mediante el procedimiento del análisis factorial confirmatorio (CFA) con ítems ordinales. Ese modelo es más apropiado para el SCC, porque a diferencia de otros sistemas de puntuación con un enfoque discreto (por ejemplo, Koppitz, 1963, 1975), cada ítem se califica en un rango de 0 a 5, y es un escalamiento de la respuesta que puede representar mejor la naturaleza continua del desempeño visomotor (Coenders & Saris, 1995; Johnson & Creech, 1983). Se usó el método máxima verosimilitud con correlaciones policóricas (Olsson, 1979) entre los ítems. Ya que las correlaciones policóricas asumen la normalidad bivariada entre los ítems (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010; Olsson, 1979), se evaluó si estas correlaciones eran apropiadas mediante índices *Root Mean Square Error of Approximation* (RMSEA; Steiger, 1990); y una prueba χ^2 de ajuste cercano para cada correlación en cada nivel de edad; estos se consiguieron con el programa LISREL (Jöreskog & Sörborm, 2006). Para la aplicación de CFA, se usó el programa EQS (Bentler, 2004).

El ajuste al modelo unidimensional se verificó considerando varios indicadores para su mejor evaluación (Hair e cols., 1998; Hu & Bentler, 1999). El estadístico de contraste fue una prueba χ^2 , que compara la matriz de covarianzas del

modelo unidimensional y el modelo nulo de completa independencia entre los ítems (Hair e cols., 1998). Como esa prueba es influenciada definitivamente por el tamaño muestral (García-Cueto, Gallo, & Mirando, 1998; Hair e cols., 1998), se aplicaron varios índices complementarios; estos fueron de ajuste absoluto (*Root Mean Square Error of Approximation*, RMSEA $\leq 0,08$; *Standardized Root Mean Residual*, SRMR $\leq 0,06$), e índices de ajuste incremental (*Comparative Fit Index*, CFI $\geq 0,95$, *Normed Fit Index*, NFI $\geq 0,90$, *Non-Normed Fit Index*, NNFI $\geq 0,90$). Se aplicó una corrección al χ^2 (χ^2 -SB, Satorra & Bentler, 1994) para atenuar el efecto de la desviación de la normalidad en el estadístico χ^2 de bondad de ajuste.

Se calcularon medidas recomendadas para evaluar la validez interna del constructo (Hair e cols., 1998) bajo el enfoque SEM: la confiabilidad del constructo (Rho; Werts, Rock, Linn, & Joreskog, 1978; Fornell & Larcker, 1981) y la varianza extraída promedio (AVE; Fornell & Larcker, 1981), que se interpretan como la convergencia de los ítems respecto al constructo latente; ambos coeficientes varían entre 0 y 1. La interpretación recomendada para aceptar como adecuados los resultados es: Rho $\geq 0,70$ y AVE $\geq 0,50$.

Como el TGB-M/SCC tiene una cobertura desde 4 años hasta 8 años y medio, los participantes con edad de 8 años 6 meses o más fueron subagrupados en un solo conjunto, denominado grupo fuera de la edad normativa (GFEN), en que la mayor edad de los participantes fue 9 años y 11 meses. Ese grupo fue separado de la muestra total para conducir los mismos análisis que la muestra total, y que permitiría una comparación entre-grupos de los efectos de la variabilidad de las respuestas sobre la dimensionalidad. La similitud entre los vectores de cargas factoriales será cuantificado por el índice ϕ de congruencia (Tucker, 1951); valores entre 0,85 y 0,94 sugieren aceptable similitud, y $> 0,95$ indicarían prácticamente igualdad entre las cargas factoriales (Lorenzo-Seva & ten Berge, 2006).

Para maximizar el poder estadístico y estabilidad de los análisis, la muestra de reagrupó en intervalos de edad que podrían representar las edades típicas en los primeros grados escolares en Perú; estos grupos fueron: Nivel 1: 4 años-0 meses a 5 años-5 meses ($n = 202$), Nivel 2: 5 años-6 meses a 6 años-5 meses ($n = 380$), Nivel 3: 6 años-6 meses a 7 años-5 meses ($n = 235$), Nivel 4: 7 años-6 meses a 8 años-5 meses ($n = 266$), y Nivel 5: 8 años-6 meses a 9 años-11 meses ($n = 286$).

Antes del análisis principal, se removieron los sujetos con puntajes extremos, debido a que pueden haber

²La ecuación para corregir la atenuación de los coeficientes alfa por restricción del rango es la siguiente (Gulliksen, 1950):

$$\alpha_x = 1 - \left[\frac{\sigma_x^2}{\sigma_{x'}^2} (1 - \alpha_x) \right]$$

σ_x^2 = varianza de la muestra restringida.
 $\sigma_{x'}^2$ = varianza de la muestra total (no restringida).
 α_x = coeficiente alfa atenuado, de la muestra restringida.
 $\alpha_{x'}$ = coeficiente alfa desatenuado o corregido.

surgido como consecuencia de la posible variabilidad de factores eventuales en la administración y en la disponibilidad motivacional del niño (Feldt & Brennan, 1989). De ese modo, se removieron los puntajes se mayores a ± 2.5 desviaciones estándares (Hair e cols., 1998).

Resultados

Análisis previo. Al evaluar el ajuste de los datos al modelo bivariado de las correlaciones policóricas, se halló que mostraban un ajuste bastante aceptable en la muestra total y en cada nivel de edad seleccionado para el presente estudio. La frecuencia modal del valor nominal p para la prueba χ^2 de ajuste fue cercana: 1,0; y la magnitud máxima y mínima para todos los niveles de

RMSEA fue 0,07 y 0,00, respectivamente. La tasa entre el resultado χ^2 de ajuste para cada correlación policórica y su grado de libertad generalmente fue menos 2,7. En resumen, las correlaciones policóricas parecen representar apropiadamente las correlaciones lineales entre los ítems.

La unidimensionalidad evaluada por el *scree test*, el análisis paralelo y el monto de varianza explicada del primer autovalor fue satisfactoriamente cumplida. Cada uno mostró inequívocamente un solo factor dominante; aunque la diferencia entre el autovalor dominante y el 2^{do} autovalor fue disminuyendo en las edades avanzadas, la diferencia entre ellas aún era sustancial. Finalmente, las distribuciones de los puntajes tuvieron ligeros desvíos respecto a la distribución normal teórica (Tabla 2).

Tabla 2
Estadísticos descriptivos y consistencia interna para los puntajes en las edades muestreadas

Edad	M	DE	Min	Max	$\sqrt{b_1}$	b_2	Consistencia interna					
							$\alpha(\alpha_{aj})$	r_{ii}	R_{ite}			
									Pro	Min	Max	
4,0m-4,5m	10,78	4,72	1	22	0,21	0,01	0,80	0,41	0,56	0,45	0,64	
4,6m-4,11m	13,26	4,05	4	23	0,29	0,53	0,76 (0,79)	0,35	0,50	0,42	0,59	
5,0m-5,5m	17,48	3,83	8	28	-0,07	-0,01	0,78 (0,83)	0,37	0,52	0,45	0,61	
5,6m-5,11m	19,11	3,76	9	28	-0,19	-0,30	0,79 (0,84)	0,38	0,53	0,49	0,57	
6,0m-6,5m	20,10	3,29	11	27	-0,34	0,06	0,73 (0,84)	0,31	0,46	0,40	0,55	
6,6m-6,11m	20,52	2,58	14	26	-0,16	-0,10	0,68 (0,89)	0,27	0,42	0,34	0,50	
7,0m-7,5m	21,41	2,42	16	28	0,10	0,01	0,67 (0,90)	0,25	0,40	0,36	0,47	
7,6m-7,11m	21,83	2,21	16	27	0,11	0,13	0,62 (0,90)	0,22	0,35	0,31	0,41	
8,0m-8,5m	23,18	2,54	17	29	0,05	-0,44	0,67 (0,89)	0,26	0,40	0,33	0,48	
8,6m-8,11m	23,50	2,49	18	29	0,17	-0,35	0,70 (0,90)	0,28	0,43	0,37	0,47	
9,0m-9,5m	23,03	2,58	15	29	-0,39	0,25	0,71 (0,90)	0,29	0,45	0,26	0,59	
9,6m-9,11m	23,31	2,75	17	28	-0,08	-0,58	0,74 (0,90)	0,33	0,49	0,28	0,66	
Edad reagrupada												
Nivel 1	14,61	4,99	1	28	-0,14	-0,21	0,84	0,48	0,63	0,58	0,65	
Nivel 2	19,54	3,58	9	28	-0,29	-0,16	0,76 (0,83)	0,35	0,51	0,48	0,57	
Nivel 3	20,99	2,53	14	28	-0,06	0,03	0,68 (0,89)	0,26	0,41	0,37	0,49	
Nivel 4	22,58	2,49	16	29	0,16	-0,23	0,67 (0,89)	0,26	0,41	0,37	0,47	
Nivel 5	23,30	2,57	15	29	-0,10	-0,11	0,71 (0,89)	0,29	0,44	0,33	0,52	
Total	20,44	4,34	1	29	-1,00	1,47	0,86	0,51	0,66	0,65	0,67	

Nota: $\sqrt{b_1}$: Simetría; b_2 : Curtosis; α : coeficiente alfa; α_{aj} : coeficiente alfa corregido por restricción del rango; r_{ii} : correlación inter-item promedio; r_{ite} = correlación ítem-test corregida.

Consistencia interna. La consistencia interna (coeficiente α) de los puntajes en cada rango de edad se reporta en la Tabla 2. Los coeficientes variaron desde 0,62 hasta 0,80, observándose que las respuestas parecen ser más consistentes en las edades más tempranas, como un efecto combinado de la correlación ítem-test entre los ítems y la dispersión del puntaje en cada nivel de edad. Es decir, que las correlaciones inter-ítem (r_{ii}) e ítem-test (r_{ite}) variaron consistentemente con la dispersión de los

puntajes. Por otro lado, el ajuste por restricción del rango (α_{aj}) aplicado a los coeficientes α resultó en coeficientes elevados, específicamente entre 0,79 y 0,80. La confiabilidad del constructo (Rho, Tabla 3), estimada de las cargas factoriales fueron menos de .80 en los tres últimos niveles de edad, pero cercanos el límite mínimo recomendado para establecer una cuantificación confiable del constructo (Rho > 0,70; Hair e cols., 1998). Por otro lado, la correlación inter-ítem (r_{ii}) mostró valores elevados en

los niveles de edad más tempranos, pero disminuyendo según el avance de las edades (Tabla 2); sus magnitudes r_{ii} pueden considerarse satisfactorios para constructos de amplia cobertura (Clark & Watson, 1995).

Análisis factorial confirmatorio. El coeficiente AVE del modelo unidimensional (Tabla 3) indica que los

primeros niveles de edad contienen un monto aceptable de varianza relacionada con el constructo latente ($> 0,50$), mientras que las demás edades se encuentran debajo de este criterio recomendado. Los autovalores (*eigen*) son congruentes con AVE, pues expresan la misma información en términos diferentes.

Tabla 3
Cargas factoriales de los ítems en la muestra total y en cada grupo de edad

Items	Total (n = 1369)		Edad 1 (n = 202)		Edad 2 (n = 380)		Edad 3 (n = 235)		Edad 4 (n = 266)		Edad 5 (n = 286)	
	λ	h^2	λ	h^2	λ	h^2	λ	h^2	λ	h^2	λ	h^2
Lámina 1	0,76	0,58	0,75	0,57	0,64	0,41	0,59	0,34	0,60	0,36	0,44	0,20
Lámina 2	0,74	0,55	0,72	0,52	0,61	0,37	0,55	0,31	0,55	0,31	0,69	0,48
Lámina 3	0,73	0,53	0,69	0,48	0,59	0,35	0,55	0,31	0,53	0,28	0,69	0,47
Lámina 4	0,76	0,57	0,75	0,56	0,63	0,40	0,69	0,48	0,54	0,29	0,60	0,36
Lámina 5	0,74	0,54	0,67	0,45	0,72	0,52	0,58	0,33	0,49	0,24	0,52	0,27
Lámina 6	0,76	0,58	0,75	0,57	0,66	0,44	0,55	0,30	0,72	0,52	0,68	0,47
Validez												
Eigen	3,54		3,4		2,77		2,33		2,30		2,47	
% var.	59,727		56,76		46,28		38,94		38,33		41,18	
AVE	0,56		0,52		0,42		0,35		0,34		0,38	
Confiabilidad												
Rho	0,88		0,87		0,81		0,76		0,75		0,78	

Nota: λ : Carga factorial; h^2 : Comunalidad. Para el agrupamiento por edad. Consultar la sección Procedimiento.

Tabla 4
Coeficientes de congruencia de Tucker entre las cargas factoriales obtenidas en cada nivel de edad

	Total	Edad 1	Edad 2	Edad 3	Edad 4	Edad 5
Total	1					
Edad 1	1,0	1				
Edad 2	0,99	0,99	1			
Edad 3	0,99	0,99	0,99	1		
Edad 4	0,99	0,99	0,99	0,98	1	
Edad 5	0,98	0,98	0,98	0,98	0,98	1

Nota: Para el agrupamiento por edad, consultar la sección Procedimiento.

La magnitud de las cargas indica una mayor varianza extraída en la muestra total, Edad 1, Edad 2 y Edad 5. La Edad 4 y Edad 5 tuvieron una mayor dispersión de las cargas y un grado de ajuste menor respecto a las otras edades y a la muestra total. Respecto a la equivalencia de las cargas factoriales entre las edades (Tabla 4). Los coeficientes ϕ de Tucker son muy elevados ($\phi > 0,97$) y superan el criterio mínimo de igualdad ($\phi > 0,95$, Lorenzo-Seva & ten Berge, 2006) entre los vectores de cargas factoriales de la muestra total y de los grupos de edad, y entre los grupos de edad.

Los resultados presentados en la Tabla 5 indican que la hipótesis de unidimensionalidad proporciona un buen ajuste con el χ^2 -SB robusto cuando es

comparado con el χ^2 obtenido por el método usual. Exceptuando la Edad 1 y 5, el χ^2 -SB de las demás edades fue estadísticamente significativo. Los residuales estandarizados (SRMR) presentados en la Tabla 5 indican un ajuste aceptable ($\leq 0,06$) de acuerdo a los niveles máximos recomendados (Hair e cols., 1998). Las magnitudes de los índices incrementales también fueron satisfactorios para la muestra total y las edades 1, 2 y 5, y relativamente bajos en la Edad 3 y 4. Los índices bajo en algunas edades siguen un patrón comparable al rango y desviación estándar de los puntajes en las edades: mejor ajuste en edades tempranas, y disminuyendo en las edades medias y un leve incremento en la última edad.

Tabla 5

Estadísticos de ajuste del modelo unidimensional; entre paréntesis y estimaciones con ajuste Saris-Bentler (SB)

	Total (n = 1369)	Edad 1 (n = 202)	Edad 2 (n = 380)	Edad 3 (n = 235)	Edad 4 (n = 266)	Edad 5 (n = 286)
χ^2 (χ^2 -SB)	122,82 (59,68)	30,65 (16,36)	58,55 (31,66)	39,75 (19,26)	40,90 (19,43)	34,18 (14,34)
χ^2 / gl^1	8,18 (3,97)	2,04 (1,09)	3,90 (2,11)	2,65 (1,28)	2,72 (1,29)	2,27 (1,02)
SRMR	0,02	0,04	0,05	0,06	0,05	0,04
RMSEA (RMSEA _{SB})	0,09 (0,06)	0,10 (0,06)	0,12 (0,08)	0,12 (0,07)	0,11 (0,06)	0,09 (0,04)
IC 90% No ajustado [Ajuste SB]	0,08, 0,11 [0,04, 0,08]	0,06, 0,15 [0,00, 0,11]	0,09, 0,15 [0,05, 0,11]	0,08, 0,16 [0,02, 0,11]	0,08, 0,15 [0,02, 0,10]	0,06, 0,13 [0,00, 0,08]
CFI	0,97 (0,98)	0,95 (0,98)	0,92 (0,96)	0,89 (0,95)	0,89 (0,95)	0,93 (0,98)
NFI	0,96 (0,98)	0,94 (0,97)	0,91 (0,94)	0,87 (0,91)	0,87 (0,91)	0,91 (0,95)
NNFI	0,95 (0,98)	0,93 (0,98)	0,87 (0,93)	0,82 (0,92)	0,83 (0,91)	0,89 (0,96)

Nota: ¹ $gl = 15$. χ^2 : prueba chi cuadrado asintótica normal. χ^2 -SB: prueba chi cuadrado escalado con el ajuste Satorra-Bentler. SRMR: Raíz media residual estandarizada. RMSEA: Raíz Media Cuadrática del Error de Aproximación. RMSEA_{SB}: RMSEA usando χ^2 -SB. CFI: índice de ajuste comparativo. NFI: Índice de ajuste normado. NNFI: Índice de ajuste no normado.

Discusión

En general, se halló que la unidimensionalidad de los puntajes para el SCC en las edades muestreadas puede considerarse satisfactoria para los fines del instrumento: el TGB-M se construyó como una evaluación de detección de niños con problemas visomotores y un complemento para la evaluación de problemas de rendimiento académico y de aprendizaje. Cuando se examinaron cercanamente los índices de ajuste, hubo un acuerdo generalizado sobre la relevancia del modelo unidimensional para explicar la variabilidad de la respuesta visomotora, pero algunos índices mostraron discrepancias. Esas discrepancias fueron notorias en algunas edades, en los que el monto de varianza retenida estuvo debajo de las recomendaciones sugeridas, es decir, debajo de 0,50 (Fornell & Larcker, 1981; Hair e cols., 1998) y debilitarían la utilidad de la interpretación válida de los puntajes. Pero resultados de ese tipo pueden ser razonables por varios motivos; primero, es difícil sostener la idea que sola la dimensión latente contenga toda la varianza de las respuestas de un examinado, especialmente en diferentes edades. Aunque los resultados respaldan que una dimensión absorbe en gran medida el desempeño del niño en el TGB-M, se podría afirmar que existen dimensiones menos dominantes que pueden explicar una parte de la variabilidad de las respuestas.

Efectivamente, durante el proceso de evaluación habrán influencias irrelevantes al constructo que, aisladamente o en interacción, producirán “ruido” en la respuesta de copiado, aún con el mejor control experimental aplicado. Hay también otras habilidades que son relevantes para la tarea de copiado (Decker, Allen, & Choca, 2006; Brannigan & Brunner, 2002), que aportan un monto de varianza sistemática y comprometen la validez de la interpretación del TGB-M como una medida pura de habilidad visomotora.

Otro de los resultados considerados satisfactorios es la similaridad de las cargas factoriales. Los coeficientes de similaridad de las cargas factoriales arrojaron muy altos valores, indicando que la magnitud de las cargas numéricas es parecida entre los diferentes grupos de edad, y que una estimación en la muestra total puede ser representativa de los grupos de edad separados.

En relación a la confiabilidad, los puntajes presentan adecuados niveles la consistencia de las respuestas. El ajuste por restricción del rango en el coeficiente α incrementó predeciblemente su magnitud, y este cambio estuvo linealmente asociado a la cuantía de la desviación estándar en cada grupo de edad. La estimación de la confiabilidad del constructo (coeficiente Rho) también convergió con el resultado del coeficiente α . Para una interpretación de los puntajes, se recomendaría usar los coeficientes de consistencia interna corregidos por restricción del rango, aunque esto puede ser discutible pues son básicamente estimaciones teóricas. La estimación de la consistencia interna por el coeficiente α supone un modelo equivalente tau en los ítems, y esta presunción no fue corroborada en el presente estudio; sin embargo, la similaridad de las cargas factoriales en cada nivel de edad (y en la muestra total, en particular) puede aproximarnos a cumplir con este presupuesto.

La discusión de la interpretación del TGB-M debe considerar también el efecto de piso en la distribución de los puntajes, que confirmaría parcialmente su limitación para aplicarse a niños de más de 8 años y 6 meses, pues Brannigan y Brunner (2002) no propusieron normas en estos grupos. Los presentes resultados no apoyan fuertemente que el TGB-M se pueda usar en niños mayores de 8 años y 6 meses, pero esta limitación debe balancearse con la finalidad de su uso. Este podría ser aún una opción válida cuando se requiere hacer un despistaje o un proxy descriptivo de la integridad

de la habilidad visomotora en niños de esas edades, sin comprometer decisiones serias. Internamente, los resultados estructurales no son pobres como para indicar una limitación completa del TGB-M en edades superiores a los 8 años y medio. De acuerdo a los presentes resultados, el TGB-M contiene un monto de varianza relevante a la visomotricidad que sería útil para determinados tipos de evaluaciones. Se requeriría confirmar esa valoración completando el estudio de la validez con el grado en que el TGB-M aporta incrementalmente la predicción de criterios relevantes (por ejemplo, rendimiento académico o discapacidades en la escritura), pues la validez de un puntaje desde el análisis factorial proporciona solo una parte de las evidencias de validez.

Dado lo anterior, los resultados del análisis estructural sugieren que el TGB-M podría ser aplicado a niños aún entre 8 años y medio y 10 años; pero tomando en cuenta la consistencia interna, la restricción del rango de sus puntajes y el monto de varianza explicada, la interpretación de sus puntajes no debe aislarse de otras fuentes de evaluación visomotora. Como medida económica, rápida, y eficiente en un contexto de carencia de instrumentos actualizados, de normas apropiadas, o de evaluaciones de grupo para propósitos de despistaje, el TGB-M/SCC puede ser un eficiente proxy para evaluar la visomotricidad entre los 8 años y medio, y 10 años. Contrariamente, no podría ser recomendada para interpretaciones clínicas.

Consideraciones finales

Durante el análisis realizado, fue posible mejorar algunos de los parámetros del modelo estudiado, como establecer errores correlacionados entre los ítems sugeridos por los índices de modificación; esto produciría una reducción de χ^2 (Brown, 2006) e incremento de los índices de ajuste, pero capitalizaría el efecto del pequeño número de ítems y el tamaño de las correlaciones entre los ítems. Eso significa que la adición de parámetros podría mejorar estadísticamente el ajuste del modelo, pero no ser sustancialmente teórica.

Los presentes resultados aportan con la validez de la estructura interna de una versión modificada del TGB, un aspecto pocas veces evaluado en la literatura científica con alguna de las versiones publicadas de esa prueba. Los resultados estructurales son satisfactorios y representan un razonable modelo para explicar la variabilidad de las respuestas a los ítems. Esas evidencias también favorecen el uso de esta versión en edades mayores a la recomendada, pero otros estudios de validez con criterios externos se requieren para dar un respaldo más fuerte a esta extensión.

Sobre las limitaciones del estudio, la generalización del estudio puede estar comprometida debido que el muestreo no se orientó a incluir explícitamente a niños de familia bilingües, y geográficamente todos los niños muestreados pertenecen a una sola provincia (Lima); es probable que el tipo de parámetros estimados no se vean afectados por las diferencias étnicas y geográficas debido que el instrumento usado es no verbal, pero no ha sido probado en el presente estudio.

Referências

- Allen, W. (1992). Book Review: The modified version of the Bender-Gestalt Test for preschool and primary school children. *Journal of Psychoeducational Assessment, 10*, 380-383.
- Arbit, J., & Zager, R. (1978). Psychometrics of a neuropsychological test battery. *Journal of Clinical Psychology, 34*(2), 460-465.
- Bender, L. A. (1938). *A visual motor gestalt test and its clinical use*. New York: American Orthopsychiatric Association.
- Merino, C., & Allen, R. A. (2012, July). A factor-analytic study for the Bender Gestalt Test, 2nd edition: Internal structure and measurement model. Paper presented in the 30th International Congress of Psychology, 22 to 27 of July, Cape Town, South Africa.
- Bentler, P. M. (2004). EQS 6.1: *Structural Equations Program [Computer program]*. Encino, CA: Multivariate Software.
- Billingslea, F. Y. (1963). The Bender Gestalt: A review and a perspective. *Psychological Bulletin, 60*, 233-251.
- Bracken, B. (2007). Creating the optimal preschool testing situation. Em: B. Bracken, & Nagle (Eds.), *Psychoeducational assessment of preschool children* (pp. 137-153). Mahwah, NJ: Lawrence Erlbaum.
- Brannigan, G. G., & Brannigan, M. J. (1995). Comparison of individual versus group administration of the Modified Version of the Bender-Gestalt Test. *Perceptual and Motor Skills, 80*, 1274.
- 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.
- Buckley, P. (1978). The Bender Gestalt Test: A review of reported research with school-age subjects, 1966-1977. *Psychology in the Schools, 15*(3), 327-338.
- Lacks, P. (1999). *Bender Gestalt screening for brain dysfunction* (2nd ed.). New York: Wiley & Sons.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. Em: R. L. Linn (Ed.), *Educational measurement* (3th ed.; pp.105-146). New York, Macmillan.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- 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.

- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245-276.
- 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.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319.
- Coenders, G., & Saris, W. E. (1995). Categorization and measurement quality: The choice between Pearson and polychoric correlations. Em: W. E. Saris, & Á. Münnich (Eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments* (pp. 125-144). Budapest: Eötvös University Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.
- Decker, S. L., Allen, R., & Choca, J. P. (2006). Construct validity of the Bender-Gestalt II: comparison with Wechsler Intelligence Scale for Children-III. *Perceptual and Motor Skills*, 102(1), 133-41.
- Decker, S. L., & Carboni, (2011). Bender Gestalt, Second Edition. Em: J. S. Kreutzer, & J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 386). New York: Springer.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing*, 18, 39-50.
- Fuller, G. B., & Vance, B. (1995). Interscorer reliability of the modified version of the Gender-Gestalt Test for preschool and primary school children. *Psychology in the Schools*, 32(4), 264-266.
- García-Cueto, E., Gallo, P., & Miranda, R. (1998). Bondad de ajuste en el análisis factorial confirmatorio. *Psichotema*, 10(3), 717-724.
- Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th Org). Hoboken, NJ: John Wiley & Sons.
- Guertin, W. H. (1952). A factor analysis of the Bender-Gestalt tests of mental patients. *Journal of Clinical Psychology*, 8, 362-367.
- Guertin, W. H. (1954a). A factor analysis of curvilinear distortions on the Bender-Gestalt. *Journal of Clinical Psychology*, 10, 12-17.
- Guertin, W. H. (1954b). A transposed factor analysis of schizophrenic performance on the Bender-Gestalt. *Journal of Clinical Psychology*, 10, 225-228.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black W. C. (1998). *Multivariate Data Analysis*. Upper Saddle River, New Jersey: Prentice.
- Haynes, J. R. (1970). Factor-analytic study of performance on the Bender-Gestalt. *Journal of Consulting and Clinical Psychology*, 34(3), 345-347.
- Holgado-Tello, F., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychromic versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153-166.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 32, 179-185.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling* 6, 1-55.
- Hutt, M. (1975). *La adaptación de Hutt del Test Gestáltico de Bender*. Buenos Aires: Guadalupe.
- International Test Commission (ITC) (2000). *Guidelines on Test Use: Spanish Version*. Translation authorized by the Colegio Oficial de Psicólogos. ITC: Author.
- Jansky, J., & de Hirsch, K. (1972). *Preventing reading failure*. New York: Harper Row.
- Johnson, D. R., & Creech, J. C. (1983.). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48, 398-407.
- Jöreskog, K., & Sörbom, D. (2006). *LISREL 8.80 for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.
- Koppitz, E. M. (1963). *The Bender-Gestalt Test for young children*. New York: Grune & Stratton.
- Koppitz, E. M. (1975). *The Bender-Gestalt Test for young children: II Research and application, 1963-1973*. New York: Grune & Stratton.
- Lee, D., Reynolds, C. R., & Willson, V. L. (2003). Standardized test administration: Why bother? *Journal of Forensic Neuropsychology*, 3, 55-81.
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's Congruence Coefficient as a meaningful index of factor similarity. *Methodology*, 2(2), 57-64.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38, 88-91.
- Merino, C. (2009). Un análisis no paramétrico de ítems de la Prueba Gestáltica del Bender Modificada para estudiantes de primaria. *Liberabit*, 15(2), 83-94.
- 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. (2011). Validez comparativa de tres sistemas de calificación para el Test Gestáltico Visomotor de Bender Modificado. *Revista de Psicología – UCV*, 13(1), 90-102.
- 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 Psychologica*, 10(1), 237-249.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Olsson, U. (1979). Maximum likelihood estimation of the polychromic correlation coefficient. *Psychometrika*, 44, 443-460.
- Parsons, L., & Weinberg, S. L. (1993). The Sugar Scoring System for the Bender Gestalt Test: An objective approach that reflects clinical judgment. *Perceptual and Motor Skills*, 77, 883-893.
- Sadeghi, R., & Hooman, H. A. (1999). A factor analysis of Bender Visual-Motor Gestalt Test. *Psychological Research*, 5(1-2), 25-39.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. Em: A. von Eye, & C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research*. Thousand Oaks, CA: Sage.
- Steiger, J. H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research*, 25, 214-12.
- Tucker, L. R. (1951). *A method for synthesis of factor analytic studies*. (Personnel Research Section Report no. 984). Washington, DC: Department of the Army.

- Walrath, R. (2011). Bender Visual Motor Gestalt Test. Em: S. Goldstein, & J. A. Naglieri (Eds.), *Encyclopedia of Child behavior and Development* (pp. 233-234). New York: Springer.
- Werts, C. E., Rock, D. A., Linn, R. L., & Joreskog, K. G. (1978). A general method of estimating the reliability of a composite. *Educational and Psychological Measurement*, 38, 933-938.

Recebido em outubro de 2012
Reformulado em fevereiro de 2013
Aprovado em abril de 2013

Sobre o autor

César Merino Soto é do Instituto de Investigación de la Universidad de San Martín de Porres.