

El sistema de calificación cualitativa para la Prueba Gestáltica de Bender – Modificada. Estudio preliminar de sus propiedades psicométricas

The qualitative scoring system for the modified version of the Bender–Gestalt Test: preliminary study of its psychometric properties

CÉSAR A. MERINO SOTO*

Abstract

The present report describes the psychometric properties of the Qualitative Scoring System for the Modified version of the Bender-Gestalt Test, in a 97 children's group, that are going to enter to first grade of elementary education in a public school of Lima. Our findings report modest levels of reliability internal consistency and test-retest reliability, in comparison with the coefficients of the manual, but their potential range (due confidence intervals) gets to high levels. The correlation with task orientation, fluid intelligence and parental report of fine-motor behavior give support to the validity and utility of this new scoring system in the context of first grade children's assessment. We discuss the limits and relevance of this scoring system for the psycho-educational research and assessment.

Keywords: Bender-Gestalt Test, children, visuo-motor integration, psychoeducational assessment, school achievement, screening.

Resumen

El presente artículo describe las propiedades psicométricas preliminares del Sistema de Calificación Cualitativa para la Prueba Gestáltica de Bender – Modificada, en un grupo de 97 niños que van a ingresar al primer grado de educación elemental en una escuela pública de Lima. Los niveles de confiabilidad por consistencia interna y test-retest fueron más modestos que los del manual, aunque su rango potencial (por intervalos de confianza) llega a niveles altos. La correlación con criterios de conducta orientada a la tarea, inteligencia fluida y el reporte parental de conductas de motricidad fina dan soporte a la validez y utilidad de este nuevo sistema en el contexto de la evaluación de niños de primer grado. Se discute su relevancia para la evaluación psicopedagógica y para la investigación educacional.

Palabras clave: Prueba Gestáltica de Bender, niños, integración visomotora, evaluación psicopedagógica, rendimiento escolar, despistaje.

* Profesor de las escuelas de Psicología de la Universidad de San Martín de Porres, de la Universidad Científica del Sur y de la Universidad Nacional Federico Villareal. Pertenece al Instituto de Investigación de Psicología de la Universidad de San Martín de Porres. Correspondencia: Calle Enrique Palacios 430, Chorrillos - Lima 9, Perú. Correo electrónico: sikayax@yahoo.com.ar.

Introducción

Como elemento significativo para la predicción del rendimiento en varias áreas del ámbito escolar, la experiencia profesional puede demostrar que la habilidad visomotriz tiende a tomar un lugar predominante en las baterías de evaluación en niños durante la edad preescolar, y en los primeros grados escolares. Por ejemplo, la prueba de coordinación motriz, en la *Batería de Molina* (Molina, 1992), luego de un análisis factorial de 15 subpruebas, se integró en el factor “estructuración espacial”, con el reconocimiento de diferencias espaciales, cierre visual y memoria visuoespacial. Además, junto con pruebas psicolingüísticas, su subescala visomotora fue incluida en la batería abreviada, dado que contribuyó a aumentar la varianza explicada en la predicción de la batería total. Molina no da mayor información psicométrica de las subpruebas, de tal modo que no se pueden saber la confiabilidad o índices de validez obtenidos en su muestra normativa, ni el grado de replicabilidad de sus parámetros psicométricos.

Paralelamente, Mora (1993) reconoció que el factor motriz es un constructo lo suficientemente diferenciado tanto en su prueba –la *Batería Evaluadora de las Habilidades Necesarias para el Aprendizaje de la Lectura y Escritura (BENHALE)*–, como en los estudios de validación de las pruebas ABC (Filho, 1960). Específicamente, la prueba de Mora exigía al niño copiar dos formas geométricas con significado: una en forma de casa y otra de cometa, las mismas que se presentaban mediante gestos. Bajo una puntuación de 0 a 4 en cada ítem, en su muestra normativa el investigador obtuvo un coeficiente de confiabilidad alfa igual a 0,72. Pero hay otras pruebas de visomotricidad que han llegado, a través de las décadas de historia de la evaluación psicológica, a posicionarse como referentes o medidas estándares de la habilidad visomotora.

La Prueba Gestáltica Visomotor de Bender (Bender, 1987) es una de ellas. Es un instrumento ampliamente utilizado para evaluar la integración visomotora, y como herramienta de despistaje para predecir el desempeño académico y la ocurrencia de problemas de aprendizaje en edades tempranas

(Sattler, 1993; Chan, 2001). Aunque Bender no desarrolló un sistema separado y validado psicométricamente, ella lo referenció como un sistema que evalúa la calidad global de cada diseño, en una escala de 1 a 5 en un diseño, y de 1 a 7 en los demás. Su sistema de calificación no ha sido extensamente usado en la práctica clínica o en la investigación, pero ha abierto un camino para una renovación de su enfoque en recientes versiones de su test (Brannigan y Brunner, 1989, 1996, 2002; Brannigan y Decker, 2003; 2006). Uno de estos sistemas será el objetivo psicométrico del presente artículo y será descrito más adelante.

Uno de los sistemas más populares para calificar las reproducciones de las figuras provenientes del Test Gestáltico de Bender es el Sistema Evolutivo de Calificación (Koppitz, 1984). Desde su publicación original en inglés, en 1964, ha sido el sistema de puntuación preferido, y se destaca porque se basa en la evaluación de errores discretos en reproducción de cada una de las 9 láminas que utiliza. Sin embargo, la evaluación con este sistema ha sido criticado, dada su sobresimplificación y el examen molecular de los errores (Brannigan y Brunner, 1989, 1996, 2002). Koppitz desarrolló el Sistema Evolutivo de Calificación consistente en 30 errores discretos, puntuados cuando están presentes en las reproducciones, se da puntuación esencialmente al error del diseño. El número de errores en cada diseño va de 2 a 4; los errores elegidos para este sistema se asumieron como predictores sensibles al rendimiento escolar, tomando en cuenta las diferencias entre estudiantes ubicados sobre o debajo del promedio. Los 30 ítems son clasificados en errores de: a) distorsión de la forma, b) rotación, c) integración de las partes y d) perseveración.

Por otro lado, hay un enfoque potencialmente útil y moderno para evaluar el desempeño en la Prueba del Bender: es el *Sistema de Calificación Cualitativa* (Brannigan y Brunner, 1989, 1996, 2002). Éste evalúa la exactitud de cada reproducción sobre una escala de 6 puntos, y en un rango de 6 a 0. Se diseñó para evaluar la calidad global de las reproducciones de los niños de 4 años y 6 meses, hasta 8 años y 5 meses. Este sistema de calificación usa un enfoque estricto de puntuación, basado en que el diseño reproducido debe ser tan

bueno o mejor que el ejemplo dado en el manual, para recibir la puntuación en el nivel de calidad correspondiente. El sistema se aplica a la versión modificada de la prueba de Bender, que únicamente utiliza 6 diseños, que considera los más apropiados en la predicción del rendimiento escolar en niños de temprana edad. El eje principal de este sistema que llama la atención de este enfoque es el abordaje gestáltico utilizado para evaluar la representación global de cada diseño copiado por el niño (Brannigan y Brunner, 2002).

Hasta la fecha, aparentemente no se ha reportado en ninguna revista arbitrada hispana algún estudio con esta nueva metodología, ni en estudios de investigación, ni en la práctica psicoeducativa profesional. El inicio de las investigaciones con este sistema debería abordar las propiedades psicométricas que respalden su uso y la elaboración de normas para su interpretación. Aunque el desarrollo de la integración visomotora tiende a estar menos afectado por otras variables culturales, como lo es la habilidad verbal (Beery, 2000), se ha reportado que los datos normativos respecto a esta área no necesariamente son constantes entre las culturas (por ejemplo, ver Chan, 2001; 2002); por lo tanto, se demanda que la elaboración de normas y los estudios psicométricos de instrumentos se estandaricen para evaluar su uso en un nuevo contexto.

El presente estudio abordará la descripción de las propiedades psicométricas de un enfoque relativamente nuevo: el *Sistema de Calificación Cualitativa* para la Prueba Gestáltica del Bender Modificada (PGBM). Este instrumento mide las habilidades de integración visomotora en el nivel preescolar y primeros grados de escuela primaria, y se utiliza como una ayuda para la identificación de problemas de aprendizaje en evaluaciones psicopedagógicas y neuropsicológicas, tal como es expuesto por los trabajos de Brannigan y Brunner en los que se ha empleado este sistema. El contexto del estudio es la evaluación psicológica profesional de niños, en varias funciones de habilidades específicas, durante su ingreso al primer grado de primaria. Por lo tan-

to, el presente estudio tiene por objetivo aportar a esta nueva versión de la PGBM, con evidencias psicométricas sobre la confiabilidad (consistencia interna y test-retest) y validez (unidimensionalidad y correlación con criterios) en una muestra de niños escolares, de la ciudad de Lima en Perú.¹

Método

Participantes

El primer grupo de participantes (n = 96) proviene de un colegio estatal ubicado en un distrito costero al sur de Lima Metropolitana. Las familias de los niños evaluados provienen, principalmente, de un nivel socioeconómico bajo a medio bajo, y de zonas urbanas y urbano-marginales. Por otro lado, la labor más frecuente entre las madres de estas familias es la hogareña. El grupo fue evaluado en dos momentos, de modo que se definieron dos grupos. Las características demográficas principales para el primer (n = 60) y segundo grupo (n = 36) fueron paralelas. La edad promedio en el primer grupo fue de 5 años y 8 meses, pero el rango de edad incluyó niños desde 5 años y 2 meses, hasta 7 años y 4 meses. En el segundo grupo, el de re-test, la edad media fue 5 años y 10 meses, y el rango total fue desde 5 años y 3 meses, hasta 7 años y 2 meses. La proporción de niños y niñas y la educación materna fueron similares en ambos grupos. La información aparece en la Tabla 1.

Tabla 1. Información demográfica de los participantes

	Grupo 1		Grupo 2. Re-test	
	N	%	n	%
Sexo del niño				
Varón	29	48,3	17	47,2
Mujer	31	51,7	19	52,8

Continúa

¹ Hasta la fecha de revisión y publicación del presente trabajo, el uso de este sistema no se ha reportado en alguna investigación con participantes hispanos, así que podría ser considerada como uno de los primeros en exponer las propiedades psicométricas de esta reciente versión de la PGBM, las mismas que incluirán hallar las evidencias de confiabilidad, validez e información normativa preliminar.

	Grupo 1		Grupo 2. Re-test	
	N	%	n	%
Edad del niño				
5a 0m - 5a 5m	6	10,0	3	8,3
5a 6m - 5a 11m	29	48,3	18	50
6a 0m - 6a 5m	20	33,3	12	33,3
6a 6m - 7a 4m	5	8,3	3	8,3
Educación materna				
Primaria (completa o inc.)	3	5	2	5,6
Secundaria completa	26	43,3	19	52,8
Secundaria incompleta	10	16,7	7	19,4
Estudios universitarios	13	21,7	5	13,9
Superior técnico	5	8,3	3	8,3
Superior técnico inc.	1	1,7	--	--
Sin dato	2	3,3	--	--
Total	60	100	36	100%

Instrumentos

Prueba Gestáltica de Bender Modificada (PGBM)

Se utilizó la PGBM, que es la versión modificada número seis de los diseños originales (A, 1, 2, 4, 6 y 8), para su aplicación en niños que cursan desde preescolar hasta los primeros grados del nivel primario (de 4,5 hasta 8,5 años), dado que son los diseños más apropiados para niños pequeños. Incluye un sistema para puntuar el desempeño gráfico del niño, el *Sistema de Calificación Cualitativa – SCC–* (Brannigan & Brunner, 2002), el cual es de 6 puntos. La puntuación de 0 a 4 se le otorga a niños que pintan líneas aleatorias y garabateo, sin concepto de los diseños, la de 5 es para aquellos cuya representación es exacta al diseño original, la de 6 es para aquello que logran gran diferenciación en la evaluación de los dibujos. Esta versión de la prueba 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 sirve como una ayuda 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 (véase, 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.

Índice de Orientación Hacia la Tarea

Los padres respondieron a 6 ítems sobre las conductas de orientación a la tarea y otras referentes a conductas de aprendizaje. Estas preguntas son parte de una escala de calificación diseñada para una identificación temprana de problemas de ajuste social (Cowen *et al.*, 1996). Los ítems fueron: 1) no hace bien abajos, 2) tiene pobres hábitos de trabajo escolar, 3) su atención es limitada, 4) tiene dificultad para seguir instrucciones, 5) tiene pobre motivación y 6) tiene dificultad para aprender. Para hacer un uso parsimonioso de esta medida en los siguientes análisis, se decidió efectuar un análisis de componentes principales para obtener un único índice representativa del constructo. El análisis de componentes reveló que un ítem (el 5), no cargaba apropiadamente con el único factor, pero sí definía un factor independiente; esto sugiere que otras influencias a nivel de constructo latente podrían explicar esta separación del ítem con el resto. Adicionalmente, en el análisis de confiabilidad se sugirió remover el ítem 5 porque no correlacionaba muy bien con el puntaje total (Rit = 0,02). Los demás ítems cargaron con valores mayores a 0,70 en el único factor. Para la interpretación de este puntaje compuesto se identificó que a mayor puntaje, el niño presentará problemas en orientación a la tarea más intensos. Su correlación con la Prueba de Inteligencia de Cattell fue bastante interesante, considerando que esta escala fue respondida por padres sin experiencia, en el llenado de escalas conductuales ($r = -.436, p < .01$).

Test de Inteligencia Libre de Cultura (Free-culture Intelligence Test)

Para la estimación de la inteligencia, se eligió la versión colectiva abreviada de esta prueba (Cattell y Cattell, 1989), que actualmente dispone de normas derivadas de la estandarización peruana en Lima Metropolitana (Altez, 1992). Está diseñada como libre de influencias culturales, lo que permite obtener una medida del factor “g”. Consta de 8 pruebas para la aplicación individual, pero la versión abreviada para la aplicación colectiva utiliza las subescalas sustitución, laberintos, identificación y semejanzas. Cada una de estas escalas tiene una prueba asignada, y la sumatoria de los resultados de todas da un puntaje total. Estas pruebas tienden a ser menos influenciadas por el trasfondo educativo y cultural de los niños, y están orientadas a medir la inteligencia fluida.

Escala de Calificación para Padres

Esta escala está basada, originalmente, en la medida aplicada por Waschbusch, Daleiden y Drabman (2000), de 24 ítems, creada para obtener información de las áreas de rendimiento cognitivo desde la teoría Horn-Cattell, para distinguir algunos aspectos del funcionamiento cognitivo (razonamiento fluido, comprensión-conocimiento, procesamiento visual, procesamiento auditivo, adquisición y recuperación), basado en la inteligencia fluida y cristalizada. Este instrumento es respondido por los padres y para ello se usa una escala de tres puntos: casi nada, moderadamente, muchas veces.

Para el presente estudio se utilizaron algunos ítems concernientes al procesamiento visual y motriz, que se eligieron bajo una observación racional de las conductas que consistentemente pueden representar aspectos visomotores del funcionamiento infantil, y cuya validez de contenido y apariencia externa se orienta a medir conductas claramente motoras y de procesamiento espacial; por ejemplo, el ítem “Tiene problemas para rasgar un papel en línea recta”. La mayoría de los ítems fueron orientados negativamente, es decir, describiendo el déficit de conducta motora. Los ítems elegidos

permitieron obtener una consistencia interna de los puntajes igual a un alfa de Cronbach de 0,79.

Procedimiento

La evaluación comprendió la aplicación de una batería de pruebas durante el proceso de admisión para el primer grado de primaria. Todas las evaluaciones se realizaron en el horario matutino. La aplicación de la PGBM formó parte de esta batería de evaluación a niños de primer grado, que comprendía instrumentos administrados a los niños y a sus apoderados, ambos de aplicación grupal. El PGBM fue la última prueba, después de otra prueba de lápiz–papel, en la que se exploraban habilidades relacionadas con el aprendizaje potencial en el primer grado de primaria, y que contenían tareas de reconocimiento de letras, percepción visual, conciencia fonológica, vocabulario y categorización.

El proceso de evaluación se inició en noviembre y terminó en febrero del año siguiente, período en que los padres estaban con expectativas de inscribir a sus niños en este colegio. La administración de las pruebas fue efectuada por dos psicólogos licenciados, quienes mantuvieron los estándares de aplicación grupal a niños (Lee, Reynolds y Willson, 2003; Bracken, 2007), y pusieron especial cuidado en que los niños no dejaran de copiar alguna figura.

Otro componente del proceso evaluativo aplicado a los participantes fue la encuesta a los padres de los niños, la cual se realizó en otra aula de clase, mediante cuestionarios respecto al desarrollo, ajuste social, satisfacción con la crianza del niño, entre estos se encontraban también los ítems descritos en la sección anterior. Cada grupo de padres de familia estuvo integrado hasta por 10 personas, monitoreadas por los examinadores.

Para la situación de test–retest, el intervalo de tiempo fue de aproximadamente dos meses y se mantuvieron similares las condiciones de administración. Sobre el procedimiento de calificación, se siguieron las instrucciones del manual; asimismo, se monitoreó el acuerdo entre el autor de este estudio y otros dos estudiantes de pregrado, que estaban aprendiendo el sistema. El nivel de acuerdo, luego de calificar 10 protocolos elegidos aleatoriamente,

fue satisfactorio, ya que las diferencias en el puntaje total entre los calificadores fue de menos de tres puntos.

Resultados

Confiabilidad

Consistencia interna

La correlación ítem-test promedio fue igual a 0,51, el cual, teniendo como referencia el mínimo recomendado (0,30, Nunnally y Bernstein, 1995), es un valor que podríamos considerar elevado en el contexto de las correlaciones entre una variable categórica (ítem) y una de intervalo (puntaje directo total). Por otro lado, la consistencia interna de los puntajes, estimada mediante el coeficiente alfa (Cronbach, 1951), ha mostrado un estatus de moderado para la muestra en el tiempo 1 y de bajo en el tiempo 2 (Tabla 2). La diferencia entre estos coeficientes se evaluó por la técnica de Feldt para muestras relacionadas (Feldt, 1980; Feldt, Woodruff y Salih, 1987), la cual fue implementada por un programa del Dr. Lautenschlager (Lautenschlager, 1989; Merino y Lautenschlager, 2003; Lautenschlager y Meade, 2008). Las diferencias entre estas estimaciones de confiabilidad no fueron estadísticamente significativas, $\chi^2(1) = 1,88$; $p = 0,16$. Dado el tamaño de la muestra, el intervalo de confianza al 95% (Onwuegbuzie & Daniel, 2001) para ambos coeficientes es amplio y llega a variar entre niveles altos y bajos de confiabilidad.

Tabla 2. Correlaciones ítem-test corregido (Rite), confiabilidad alfa de Cronbach y estadísticos del análisis de componentes principales

	Total (n = 60)		
	Rite	Cargas	h ²
BGA	0,67	0,81	0,67
BG1	0,47	0,65	0,42
BG2	0,38	0,52	0,27
BG4	0,44	0,60	0,36
BG6	0,61	0,76	0,59

Continúa

	Total (n = 60)		
	Rite	Cargas	h ²
BG8	0,54	0,71	0,51
Alfa de Cronbach	0,77 [0,66 - 0,85] ^a		
	0,62 [0,38 - 0,78] ^b		

A = Alfa de Cronbach en el tiempo 1

b = Alfa de Cronbach en el tiempo 2 (n = 36)

Estabilidad test-retest

La correlación test-retest indica una moderada estabilidad ($r = 0,503$; $n = 36$; $p < 0,01$). Coeficientes de magnitud similar, aunque algo mayores, se han hallado en pruebas de integración visomotriz que consideran el intervalo de tiempo muestreado, por ejemplo 7 meses, en niños preescolares (Beery, 2000).

Validez

Unidimensionalidad

Para estimar este aspecto, se usó un análisis factorial común, con el objetivo de obtener la asociación multivariada entre los ítems y la dimensión subyacente a ellos. Al hacer un análisis de autovalores de la matriz correlacional inter-ítem, más la regla de Kaiser (Nunnally y Bernstein, 1995), sólo un factor obtuvo un *eigenvalue* mayor a 1 (2,84), explicando el 47% de la varianza total; su equivalente gráfico en el *screen test* (Cattell, 1966) inequívocamente también mostró que un solo factor se puede extraer confiablemente de los datos. Ya que se obtuvo un valor de varianza superior al 20%, explicada para poder asumir la unidimensionalidad de la medida bajo estudio (Reckase, 1979), podemos afirmar que la unidimensionalidad que subyace a las relaciones entre los ítems. Adicionalmente, las cargas factoriales extraídas mediante un análisis factorial exploratorio mostraron que los ítems tuvieron magnitudes sustanciales ($f > 0,50$) sobre el único factor latente (Tabla 2).

Correlación con criterios

En la Tabla 3 se muestran las correlaciones lineales entre las medidas de criterio y el puntaje directo del PGBM. La correlación del PGBM con el CI del *Test de Inteligencia Libre de Cultura*, una medida de inteligencia fluida, ha sido moderada-alta, mientras que con el índice de orientación hacia la tarea, la correlación ha sido moderada-baja y negativa ($r = -0,34$; $p < 0,05$) después de controlar la influencia de la inteligencia del niño ($r = -0,51$; $p < 0,01$). Por otro lado, las correlaciones con la otra escala para padres, respecto a habilidades cognitivas, han dado 6 de 13 correlaciones significativamente estadísticas, pero de baja magnitud ($r < 0,25$). Éstas han relacionado el PGBM con conductas más asociadas al manejo de la mano para tareas de lápiz-papel (como dibujar), las cuales podrían requerir de un control más preciso, para rasgar sobre una línea recta o para manejar bloques. Interesantemente, dos ítems que podrían estar asociados a signos de deficiencia en la capacidad visomotriz también han relacionado: borrones ($r = -0,15$) y dolor de mano por el uso del lápiz ($r = -0,23$).

Tabla 3. Correlaciones de la Prueba de Bender Modificada con inteligencia y orientación a la tarea

Criterio	R Pearson
CI del test de Cattell	0,50**
Task Orientation Index	
Correlación de orden cero	-0,51**
Correlación parcial (controlando CI)	-0,34*
Escala de Calificación de Habilidades – Padres	
E2 Tijeras	-0,09
E5 Línea recta	-0,21**
E6 Rasgar recto	-0,14*
E13 Bloques	-0,14*
E19 Copia lento	-0,12
E32 Borrones	-0,15*
E34 Movimiento de manos	-0,21**
E38 Demora dibujo	-0,12
	Continúa

Criterio	R Pearson
E41 Imitar movimientos	0,007
E46 Detalle de lo que ve	-0,14
E53 Límite de tiempo	-0,09
E55 Dolor en mano por lápiz	-0,23**
E56 Repaso, pintar	-0,118

** = $p < 0,01$

* = $p < 0,05$

Información normativa preliminar

El desempeño de los niños del presente estudio fue similar entre ambas submuestras (Tabla 4). De este modo, el rendimiento medio del puntaje directo para el primer grupo fue igual a 20,45 (d.e. = 3,48); y para el segundo grupo fue igual a 21,58 (d.e. = 2,1), lo que revela diferencias leves entre ellos. Por su parte, el puntaje directo varió desde 20,5 (en niños de 6 a 6 años y medio) hasta 22,02 (en niños de 6,5 años a 7 años y 4 meses). Este desempeño en la muestra total es muy similar al desempeño promedio de niños norteamericanos entre 7 y 7,5 años (Brannigan y Brunner, 2002) y prácticamente igual al desempeño de niños de Hong Kong entre 5 años y medio y 5 años y 11 meses (Chan, 2001; 2002).

Por otro lado, las Pruebas de Normalidad Shapiro – Wilk (Shapiro y Wilk, 1965) para la muestra 1 dieron como resultado 0,947 ($g.l. = 36$; $p = 0,08$), y para la muestra re-test, 0,977 ($g.l. = 36$; $p = 0,64$), lo que indica que la distribución de los puntajes no es distinta a la de tipo normal.

Tabla 4. Información normativa preliminar: Prueba Gestáltica de Bender – Modificada, Sistema de Calificación Cualitativa

	Evaluación tiempo 1		Evaluación tiempo 2	
	M	DE	M	DE
5a 0m - 5a 5m	21,50	3,93	20,33	2,08
5a 6m - 5a 11m	20,48	3,54	22,11	2,39
6a 0m - 6a 5m	19,65	3,51	21,00	1,70
6a 6m - 7a 4m	22,20	2,16	22,00	1,00
Total	20,45	3,48	21,58	2,10

Discusión

Aunque desde la creación de la Prueba Gestáltico Visomotor de Bender se han generado muchas investigaciones sobre sus correlatos, neuropsicológicos y clínicos, con el desempeño escolar, así como sobre nuestras formas de calificación, generalmente éstas no han circulado en el mundo hispano. Una de estas variaciones modernas es el Sistema de Calificación Cualitativa para la versión modificada de la Prueba Gestáltica de Bender (Brannigan y Brunner, 1989, 1996, 2002). Debido a que no hay información psicométrica de esta modificación en el mundo hispano y que es una potencial y útil medida para la práctica evaluativa y la investigación aplicada, el presente estudio tuvo por finalidad exponer las evidencias iniciales de validez y confiabilidad en una muestra de participantes de la ciudad de Lima, en Perú. Las diferencias en la magnitud de las correlaciones han tenido una esperable variabilidad, aun considerando que espúreamente se han visto influenciadas por el rango de los puntajes en cada variable criterio; esto quiere decir que el puntaje de CI tiene un amplio rango de posibles valores, mientras que el rango para el puntaje compuesto de orientación hacia la tarea es menor; y, mucho menor en cada ítem de la escala de habilidades para padres (4 opciones).

Aún con ello, las correlaciones para explorar la validez de constructo han sido moderadas y en la dirección de estudios previos. Por ejemplo, la correlación inteligencia e integración visomotriz comparte cerca del 25% de la varianza respecto al uso de la Prueba de Desarrollo de la Integración Visomotora (Beery, 2000). Soledad, Maganto y Garaigordobil (2001) reportan similar grado de varianza entre estos constructos.

Las correlaciones con este criterio también sugieren que fueron los ítems cuyas tareas requerían la destreza digital sobre el papel y dentro del contexto del escritorio (dibujar línea recta, dolor en mano por lápiz, y movimiento de manos). Aunque estas correlaciones son bajas, teóricamente aún señalan la relevancia de la función de integración visomotriz para el desempeño de tareas académicas cotidianas y alguna sintomatología de incoordinación ojo-mano. Ello está en concordancia con indi-

cadores en el aula de clase sobre los déficits de procesamiento psicológico básicos en la integración visomotriz: dificultad para copiar de la pizarra, copiar diseños y dificultades con el trazado de líneas y agarre del lápiz, que a la vez indican particulares problemas en el desarrollo de la simultaneidad, alternancia y disociación del movimiento (Colarusso y O'Rourke, 1999; Berdicenski y Milicic, 2002). Es de notar que el nivel de exactitud del reporte de los padres también es un factor que puede haber atenuado las correlaciones convergentes entre las reproducciones del Bender y los ítems del cuestionario; sin embargo, la información proveniente de los padres puede lograr grados muy aceptables para la investigación y para el despistaje de problemas en los niños (Glascoe, 2002).

Las diferencias normativas entre la presente muestra de estudio y las del manual han sugerido que el contexto cultural y los eventos instruccionales formales e informales han impactado, posiblemente, en el patrón de habilidades de integración visomotriz. Aunque estas diferencias pueden provenir de la varianza del calificador de las reproducciones (examinador), las diferencias con la muestra normativa norteamericana también han ocurrido en niños asiáticos (Chan, 2001; 2002). Estas diferencias entre la muestra piloto peruana y las reportadas en el manual Bender (Brannigan y Brunner, 2002) y en la obra de Chan (2001, 2002) deben llevar a una posterior investigación normativa para replicar los presentes resultados. Los efectos de la cultura en el desarrollo y aprendizajes para resolver tareas cognitivas y visomotoras han estado vinculados con patrones de diferentes oportunidades para aprender en contextos formales e informales (Rosenblum, Katz, Hahn-Markowitz, Mazor-Karsenty y Parush, 2000; Katz, Kinozy y Parush, 2001).

El valor de la prueba modificada del Bender está delimitado por el sistema de puntuación, de tal modo que los criterios e instrucciones para calificar las reproducciones de los niños son el puente para estimar el desempeño visomotor. Por lo tanto, las diferencias en el niño pueden deberse a las diferencias en la calificación del examinador, la misma que corresponde a la observación molar y gestáltica de las reproducciones gráficas de los niños. Justamente, estudios de confiabilidad interjueces son

necesarios y el realizarlos en la muestra constituye un objetivo del investigador del presente artículo.

El método presentado aquí, el Sistema de Calificación Cualitativa (Brannigan y Brunner, 2002), difiere del popular sistema de Koppitz principalmente en su brevedad, en su moderadamente amplio rango de puntuación para cada figura y en el enfoque molar. Actualmente, este método no pone ningún énfasis o sugiere interpretaciones emocionales o de ajuste social, más allá de los declarados por el manual sobre el rendimiento visomotor. Tal aspecto merece ser enfatizado porque largamente se ha sustentado que las pruebas de copiado o dibujo proporcionan un índice de la inteligencia o el funcionamiento emocional del niño, aunque hay buenas razones para dudar de estas afirmaciones (Simner, 1983).

Una ventaja adicional del método aquí reportado es su posibilidad de aplicación en grupo en situaciones de evaluación de despistaje, ya que ha sido comprobado que este método grupal no afecta seriamente la variabilidad de los resultados comparándolo con la aplicación individual (Brannigan & Brunner, 2002). Aparentemente, la aplicación grupal de tareas de integración visomotora no es un factor de serio peligro para la validez de los puntajes obtenibles, situación que ha sido comprobada anteriormente (Pryzwansky, 1977; Beery, 2000)

Hay más pasos adicionales que se deben seguir para probar la utilidad de este método en la predicción del rendimiento escolar en investigación y su uso profesional en instituciones de educación básica, considerando que el uso de pruebas sensorio-motoras relativamente nuevas, aparentemente, está llamando la atención de los investigadores y profesionales (Parush, Rilsky, Goldstand, Mazor-Karsenty y Yochman, 2002). En este sentido, las baterías aplicadas en la evaluación de niños que ingresan a primer grado pueden hallar que un sistema relativamente novedoso para el medio latinoamericano puede fortalecer la práctica profesional y principalmente la investigación predictiva, tomando en cuenta que las notas en el primer grado en lectura y aritmética tienden a ser poderosos predictores del desempeño escolar en la educación secundaria (Simner y Barnes, 1991). También debemos señalar que los problemas de aprendizaje no-verbales

pueden ser advertidos por un bajo rendimiento en el Bender, considerando que un signo típico de este déficit es la discrepancia confiable del rendimiento de la cognición espacial frente a otras medidas cognitivas verbales (Pennington, 1991). La validez predictiva del PGBM usando este sistema de calificación cualitativa podría mejorar las sensibilidades de las diferencias entre niños con escritura normal o pobre, considerando que los problemas en la escritura manual requieren de la precisión del copiado que también se pone en juego durante la reproducción de las figuras de la PGBM. Mientras el interés por los niños y su desarrollo continúe en el ojo de la investigación sobre las herramientas que ayudan a detectar sus problemas, el impulso hacia nuevos métodos como el aquí descrito también continuará en la agenda de los investigadores educacionales.

La presente investigación no es conclusiva sobre los hallazgos reportados, y se recomienda expandir las evidencias de validez del Sistema Cualitativo de Calificación de Brannigan y Brunner (2002), así como evaluar su utilidad predictiva y descriptiva en la evaluación de niños. Con tal información, este sistema puede ofrecer un interesante y valioso medio de evaluación, el cual competiría con el antiguo y popular método de Koppitz (1984), no únicamente en el contexto geográfico en que se realizó la investigación (Perú), sino también en Latinoamérica.

Referencias

- Altez, I. (1992). *El test de inteligencia, Factor G, Escala I, Forma abreviada colectiva de R.B. Cattell estandarizado en Lima Metropolitana*. Lima: Universidad Femenina del Sagrado Corazón.
- Beery, K.E. (2000). *Prueba Beery-Buktenica del Desarrollo de la Integración Visomotriz* (4ta ed.). México, D.F.: El Manual Moderno.
- Bender, L. (1987). *El Test Gestáltico Visomotor*. Buenos Aires: Paidós.
- Berdicewski, O. y Milicic, N. (2002) *Manual de prueba de funciones básicas*. Chile: Universtaria.
- Bracken, B. (2007). Creating the optimal preschool testing situation. In: B. Bracken & Nagle (eds.). *Psychoeducational assessment of preschool children* (pp. 137-153). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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., Decker, S. L. y Madsen, D. H. (2004). Innovative features of the Bender-Gestalt II and expanded guidelines for the use of the Global Scoring System. In: G. G. Brannigan, S. L. Decker & D. H. Madsen. *Bender Visual-Motor Gestalt Test* (2da. ed.) (pp.). Ithaca, IL: Riverside Publishing.
- Brannigan, G.G. y Decker, S.L. (2003). *Bender Visual-Motor Gestalt Test* (2da. ed.). Ithaca, IL: Riverside Publishing.
- Brannigan, G.G. y Decker, S.L. (2006). The Bender-Gestalt II. *American Journal of Orthopsychiatry*, 76, 10-12.
- Cattell, R. B. y Cattell, A. K. S. (1989). *Test de Factor "g", escala I*. Madrid: TEA.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 629-637.
- Chan, P. W. (2001). Comparison of visual motor development in Hong Kong and the USA assessed on the Qualitative Scoring System for the Modified Bender-Gestalt Test. *Psychological Reports*, 88, 236-240.
- Chan, P. W. (2002) Relationship of the visual motor development and academic performance of young children in Hong Kong assessed on the Bender-gestalt test. *Perceptual and Motor Skills*, 90, 209-214.
- Colarusso, R. y O'Rourke, C. (1999). *Special education for all teachers* (2da. ed.). Dubuque, IA: Kendall/Hunt.
- Cowen, E. M., Hightower, A. D., Pedro-Carroll, J. L., Work, W. C., Wyman, P. A. & Haffey, W. G. (1996). *School-based prevention for children at risk. The Primary Mental Health Project*. Washington, DC: American Psychological Association.
- Filho, L. (1960). *Test ABC de verificación de la madurez necesaria para el aprendizaje de la lectura y escritura* (6ta. ed.). Buenos Aires: Kapelusz.
- Feldt, L. S. (1980). A test of the hypothesis that Cronbach alpha-reliability coefficient is the same for 2 tests administered to the same sample. *Psychometrika*, 45, 99-105.
- Feldt, L. S., Woodruff, D. J. & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11(1), 93-103.
- Glascoe, F. P. (2002). *Collaborating with parents*. Antioch, Tennessee: Ellsworth & Vandermeer Press.
- Katz, N., Kizony R. & Parush, S. (2001). Visuomotor organization and thinking operations performance of school age ethiopian, bedouin and mainstream israeli children. *Occupational Therapy Journal of Research*, 22(1), 34-43.
- Köppitz, E. M. (1984). *El test gestáltico visomotor para niños* (10ma. ed.). Buenos Aires: Guadalupe.
- Lautenschlager, G. J. (1989). AlphaTest. Testing for differences in values of coefficient alpha. *Applied Psychological Measurement*, 13, 284.
- Lautenschlager, G. J. & Meade, A. W. (2008). AlphaTest. A window's program for tests of hypotheses about coefficient alpha. *Applied Psychological Measurement*, 32, 502-503.
- Lee, D., Reynolds, C. R. & Willson, V. L. (2003). Standardized test administration: why bother? *Journal of Forensic Neuropsychology*, 3, 55-81.

- Merino, C. y Lautenschlager, G. (2003) Comparación estadística de la confiabilidad alfa de Cronbach: Aplicaciones en la medición educacional y psicológica. *Revista de Psicología de la Universidad de Chile*, 12(2), 127-136.
- Molina, S. (1992). *BADICBALE: Bateria Diagnóstica de la Competencia Lectora Básica para el Aprendizaje de la Lectura*. Madrid: CEPE.
- Mora, J. (1993). *BEHNALE: Bateria Evaluadora de las Habilidades Necesarias para el Aprendizaje de la Lectura y Escritura*. Madrid: TEA.
- Nunnally, J.C. y Bernstein, I.J. (1995). *Teoría psicométrica* (3ra ed.). México, D. F.: McGraw-Hill.
- Onwuegbuzie, A. y Daniel, L.G. (2001). Indexes of score reliability and their applications. Ponencia expuesta en la reunión anual de la American Educational Research Association, Abril 10-14, Seattle, Washington.
- Parush, S., Rilsky, A., Goldstand, S., Mazor-Karsenty, T. y Yochman, A. (2002). The use of the QNST-II for the identification of children with perceptual-motor deficits. *Occupational Therapy International*, 9(3), 185-200.
- Pennington, B. F. (1999). *Diagnosing learning disorders. A neuropsychological framework*. New York: Guilford.
- Pryzwansky, W. B. (1977). The use of the developmental test of visual-motor integration as a group screening instrument. *Psychology in the Schools*, 14(4), 419- 423.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results implications. *Journal of Educational Statistics*, 4, 207-230.
- Rosenblum, S., Katz, N., Hahn-Markowitz, J., Mazor-Karsenty, T. y Parush, S. (2000). Environmental Influences on perceptual and motor skills of children from immigrant ethiopian families. *Perceptual and Motor Skills*, 90, 587-594.
- Sattler, J. (1993). *Evaluación de la inteligencia infantil y habilidades especiales* (2da. ed.). México, D.F.: El Manual Moderno.
- Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 5, 591-611.
- 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.
- Simner, M.L. (1989). Predictive validity of an abbreviated version of the Printing Performance School Readiness Test. *Journal of School Psychology*, 27, 189-195.
- Soledad, M., Maganto, C. y Garaigordobil, M. (2001). Análisis evolutivo de la coordinación visomotora y sus relaciones con inteligencia, estilo cognitivo y atención. *Revista Española de Orientación y Psicopedagogía*, 12(21), 212-222.

Recepción: 11 de octubre de 2008
Aceptación: 18 de mayo de 2009