

Validez incremental del Test Gestáltico de Bender Modificado, en niños que inician el primer grado

Validade incremental do Teste Gestáltico de Bender Modificado, em crianças começando a 1ª série do Ensino Fundamental

Incremental Validity for the Modified Bender Gestalt Test in Children Commencing First Grade

César A. Merino Soto*

Universidad de San Martín de Porres, Perú

Doi: [dx.doi.org/10.12804/apl32.2.2014.07](https://doi.org/10.12804/apl32.2.2014.07)

Resumen

El Test Gestáltico de Bender (TGB) ha tenido una larga historia en la evaluación psicológica y se han desarrollado varios sistemas de calificación. Uno de ellos es el Sistema de Calificación Cualitativa (SCC) de Brannigan y Brunner, desarrollado para la versión modificada de esta prueba que usa seis de los nueve diseños originales (TGB Modificado). Aún hay pocas evidencias psicométricas en habla hispana sobre la utilidad de esta prueba en una batería de despistaje académico. El propósito de este estudio es reportar evidencias de validez incremental de esta versión abreviada, respecto al rendimiento escolar, en un grupo de 101 niños de 5 años en etapa de ingreso al primer grado de educación primaria. Se aplicó una regresión múltiple jerárquica controlando los efectos de la inteligencia fluida y las habilidades pre-académicas, frente al criterio de las notas escolares. Se halló que la inclusión del TGB Modificado aporta significativa varianza única en las notas escolares, y que es un predictor más potente que la inteligencia. Nuestros resultados respaldan el uso del TGB Modificado en una batería breve de despistaje en la identificación de niños en riesgo de bajo rendimiento escolar.

Palabras clave: Test Gestáltico de Bender, validez, evaluación, niños, psicometría, sistema cualitativo

Resumo

O Teste Gestáltico de Bender (TGB) tem tido uma longa história na avaliação psicológica, e têm se desenvolvido vários sistemas de qualificação. Um deles é o Sistema de Qualificação Qualitativa (SQQ) de Brannigan e Brunner, desenvolvido para a versão modificada desta prova que usa seis dos nove modelos originais (TGB Modificado). Ainda existem poucas evidências psicométricas em língua espanhola sobre a utilidade desta prova em uma bateria de blindagem acadêmica. O propósito deste estudo é reportar evidências de validade incremental desta versão abreviada respeito ao rendimento escolar, em um grupo de 101 crianças de cinco anos, em etapa de começar a 1ª série do Ensino Fundamental. Aplicou-se uma regressão múltipla hierárquica controlando os efeitos da inteligência fluída e as habilidades pré-acadêmicas, frente ao critério das notas escolares. Encontramos que a inclusão do TGB Modificado aporta significativa variância única nas notas escolares, e que é um predictor mais potente que a inteligência. Nossos resultados apoiam o uso do TGB Modificado em uma bateria breve de blindagem na identificação de crianças em risco de baixo rendimento escolar.

* César A. Merino Soto, Instituto de Investigación, Escuela de Psicología, Universidad de San Martín de Porres. La correspondencia relacionada con este artículo debe ser enviada a César A. Merino Soto. Correo electrónico: sikayax@yahoo.com.ar

Para citar este artículo: Merino, S. C. A. (2014). Validez incremental del Test Gestáltico de Bender Modificado, en niños iniciando que inician el primer grado. *Avances en Psicología Latinoamericana*, 32(2), 275-286. doi: [dx.doi.org/10.12804/apl32.2.2014.07](https://doi.org/10.12804/apl32.2.2014.07)

Palavras-chave: Teste Gestáltico de Bender, validade, avaliação, crianças, psicometria, sistema qualitativo

Abstract

The Bender Gestalt Test (BGT) has a long history in psychological assessment and several scoring systems have been developed for it. One of them is the Brannigan and Brunner's Qualitative Scoring System (QSS), developed for the modified version of this test that uses six of the nine original designs (BGT Modified). However, there is limited psychometric evidence for the Spanish speaking population about the usefulness of this measure in an academic screening battery. The purpose of this study is to report evidence of incremental validity of this abbreviated version regarding school achievement, within a group of 101 five year olds entering the first grade of primary education. We applied a hierarchical multiple regression controlling for the effects of fluid intelligence and pre-academic skills, compared to the criteria of school grades. We found that the inclusion of BGT Modified brings significant unique variance in school grades, and is a more powerful predictor than intelligence. Our results support the use of BGT Modified in a brief battery of screening in identifying children at risk for poor school performance.

Key words: Bender Gestalt Test, validity, assessment, children, psychometry, qualitative system

Demostrar si una nueva medida es relevante a un constructo de interés para explicar una varianza significativa de un criterio es un aspecto de la investigación sobre la validez incremental de una nueva medida. La validez incremental consiste en el grado en que una nueva medida puede ser capaz de mejorar la predicción de un criterio en el contexto de otras medidas ya existentes (Haynes & Lench, 2003; Husley & Meyer, 2003). Este mejoramiento significa aumentar la información práctica y psicométrica más allá de lo aportado por los instrumentos existentes (Haynes & Lench, 2003; Haynes, Nelson & Blaine, 1999), y que adiciona información sobre la contribución práctica o estadística de un nuevo instrumento. Desde la concepción de este aspecto de la validez se intenta hallar

evidencias que proporcionen información relevante a la utilidad de un instrumento, para añadir información adicional en la explicación de un criterio en el contexto de otros predictores o instrumentos ya existentes (Schrest, 1963).

Esto se puede ejemplificar mejor si tomamos en cuenta que en el contexto escolar las evaluaciones escolares psicológicas consisten en una variada combinación de medidas que buscan describir lo más completamente posible al estudiante que inicia el año escolar. La evaluación de despistaje, fundamentada en la detección de riesgos conductuales que pueden influir en la adaptación y rendimiento académico futuro (Salvia & Ysseldyke, 1985), es una de estas evaluaciones características del trabajo del psicólogo en las instituciones educativas básicas (Merino, Díaz, Zapata & Benites, 2006). Para tal fin, los psicólogos escolares cuentan con herramientas estandarizadas o creadas por ellos mismos (Mercer, 1991), pero el advenimiento de nuevos instrumentos lleva al psicólogo a preguntarse sobre su valor pragmático en la identificación de futuros problemas en los estudiantes. Por lo tanto, el profesional debe considerar un número de criterios para lograr un buen balance de los beneficios y los costos de usar tal instrumento. Esto puede significar ver el costo monetario del instrumento, el tiempo de aprendizaje y de administración y obtención de las puntuaciones, la información psicométrica relevante, el grado de comunicabilidad de los resultados y el modo en que mejora la predicción de un criterio (por ejemplo, rendimiento académico). Además de los criterios mencionados, se puede estimar cuantitativamente el grado en que la nueva medida incrementa la validez predictiva de los instrumentos ya existentes, y si esta varianza única del nuevo instrumento es mínimamente suficiente como para incluirla en la batería de pruebas ya existentes. Esta situación parece común en las evaluaciones de niños que ingresan al primer grado, en el que se busca detectar a niños con potenciales problemas en el aprendizaje futuro de la lectura y otras materiales escolares. Por lo tanto, la evaluación de la validez incremental de un instrumento de medición se caracteriza por ser multidimensional (Johnston & Murray, 2003) y no únicamente examinar criterios prácticos o estadísticos independientemente.

Considerando que una de las habilidades de mayor impacto durante los primeros años de escolaridad es la habilidad visomotora (Aronholt, Fraas & Fraas, 1990; Beery, 2000; Koppitz, 1984; Kulps, 1999; Sortor & Kulp, 2003), se han creado varias medidas de este constructo. Por ejemplo, el Test Gestáltico de Bender (TGB; Bender, 1938) se creó para esta finalidad y parece ser una de las herramientas populares incluidas en las evaluaciones para niños (Archer & Newsom, 2000; Piotrowski, 1995; Sullivan & Bowden, 1997). Actualmente, se han publicado nuevas versiones con cambios importantes en la estructura y calificación (por ejemplo, Brannigan & Decker, 2003; Reynolds, 2007). Retrospectivamente, las modificaciones más frecuentes ocurrieron en el sistema de calificación de los dibujos reproducidos y, sin duda alguna, una de las versiones internacionalmente más usadas fue el Sistema Evolutivo de Koppitz (1984), que se basa en una evaluación discreta de los errores cometidos por el niño en las reproducciones. En contraste con el sistema de Koppitz, posteriormente se crearon métodos basados en la observación holista para calificar los dibujos reproducidos del BGT, una de ellas fue el Sistema Cualitativo de Calificación (SCC; Brannigan & Brunner, 2002), creada a partir de una reducción del TGB original (de nueve a seis láminas: TGB Modificado, TGB-M).

En general, las revisiones tempranas sobre el papel del TGB en la predicción del rendimiento educativo (Billingslea, 1963; Keogh & Smith, 1961, 1969; Norfleet, 1973) respaldan su capacidad predictiva al explicar parte de la varianza de este criterio, así como en la detección temprana de problemas de problemas de aprendizaje (Telegdy, 1974a, 1974b, 1975; Wallbrown, Engin, Wallbrown & Blaha, 1975; Wallbrown, Wallbrown & Engin, 1974). Sin embargo, otra parte de la investigación aplicada ha mostrado solo una modesta e inconclusa predictividad en el rendimiento escolar (Buckley, 1977; Coy 1974). Desde hace tiempo existen algunas evidencias que reportan la limitada utilidad de las pruebas que poseen un fuerte componente maduracional, como el VMI o el Bender (Colarusso, Gill, Plankenhorn & Brooks, 1980 y Larsen & Hammill, 1975). Debe tomarse nota que estas

investigaciones se refieren al TGB adaptado por Koppitz (1984).

La presente investigación evaluará el valor predictivo del TGB-M respecto a las calificaciones escolares en niños de primer grado de primaria, mediante la aplicación de un enfoque de validez incremental del TGB-M, que incorpora el análisis de regresión lineal jerárquico para aislar la varianza específica del instrumento que nos interesa del resto de las otras variables predictoras. Esto se efectuará al introducir paso a paso las variables explicativas para observar el incremento en la varianza explicada sobre la variable dependiente (en esta investigación, las calificaciones escolares). Se asumirá como punto de partida que la correlación lineal entre el TGB-M y rendimiento escolar puede contener una proporción de varianza común con otra variable (por ejemplo, inteligencia), que explica parte de la covariación entre ellas. Si el TGB-M puede adicionar una cantidad de varianza a los predictores existentes sobre el rendimiento académico, entonces su contribución debe evaluarse desde su significancia pragmática y estadística. Considerando el número de variables que se incluirán como predictores, es posible que una tercera variable produzca incrementos pequeños estadísticamente significativos (Nunally & Bernstein, 1995). Hasta la fecha, en América Latina no hay investigaciones empíricas sobre la capacidad predictiva del TGB-M respecto al rendimiento escolar, y por lo tanto la presente investigación pretende dar un soporte a la validez psicométrica de esta versión del TGB.

Método

Participantes

La población del presente estudio son los niños y niñas en etapa de las evaluaciones de ingreso al primer grado de primaria en colegios estatales de un distrito costero al sur de Lima Metropolitana. Los niños egresaron recientemente de sus estudios preescolares en instituciones educativas alrededor del lugar de recolección de datos. La población estudiada pertenece a una de las jurisdicciones administrativas del Ministerio de Educación del Perú, llamada UGEL (Unidad de Gestión Educativa

Local), registro 07. Este distrito 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). El distrito contiene esencialmente áreas urbanas y sub-urbanas.

Los participantes de la muestra fueron 106 niños (varones, $n = 60$, 56.6%), de una edad promedio de 69.4 meses ($d.e. = 4.43$ meses; mín. = 59 meses; máx. = 78 meses), todos provenientes de 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 funcionales, estructurales y organizacionales de estas instituciones educativas preescolares son similares en líneas generales entre ellas, y por lo tanto 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. La habilidad cognitiva general se distribuyó levemente con asimetría negativa (coeficiente de Fisher = -0.39), según los resultados de una prueba de inteligencia (ver la sección Instrumentos). Dado la zona de ubicación de los colegios, racionalmente se puede afirmar que las familias de los niños típicamente alcanzan mayormente el nivel de instrucción secundario básico, y las madres tienden a pasar más horas con el niño, ya que se ocupan del hogar y eventualmente realizan actividades económicas independientes. 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 inferior, perciben alrededor de 500 dólares americanos mensuales o menos.

Instrumentos

Test Gestáltico de Bender Modificado (Brannigan & Brunner, 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. Esta versión,

así como su sistema de puntuación, proviene del trabajo de DeHirsh (DeHirsh et ál., 1966; Jansky & deHirsh, 1972) y Keogh (Keogh & Smith, 1961). Incluye un sistema para puntuar el desempeño gráfico del niño, el *Sistema de Calificación Cualitativa (SCC)* (Brannigan & Brunner, 2002) de seis puntos, desde una puntuación de cero (líneas aleatorias, garabateo, sin concepto de los diseños) hasta cinco (representación exacta del diseño); y que logra 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 dibujos 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 para problemas de aprendizaje y descripción del desarrollo visomotor del niño (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). La consistencia interna de la puntuación del *SCC* en nuestra muestra fue $\alpha = 0.77$, mientras que la correlación intraclase entre el autor del presente estudio y un colaborador que puntuaron los diseños reproducidos de 25 niños elegidos aleatoriamente de la muestra, fue 0.80.

Test de Inteligencia Libre de Cultura de Cattell (Cattell & Cattell, 1989; Altez, 1992). Para la estimación de la inteligencia, se eligió la versión colectiva abreviada de esta prueba (Cattell & Cattell, 1989), que actualmente dispone de normas derivadas de la estandarización peruana en Lima Metropolitana (Altez, 1992). Está diseñada como una medida cuyos estímulos minimizan las influencias culturales en su elaboración, lo que permite obtener una estimación del factor “g” o la llamada inteligencia fluida. Consta de ocho pruebas para la aplicación individual, pero la versión abreviada para la aplicación colectiva utiliza las subpruebas Sustitución, Laberintos, Identificación y Semejanzas, que son las que mejores condiciones que se presentan para aplicaciones de despistaje de la habilidad intelectual. El instrumento arroja

una sola puntuación derivada de la suma de las cuatro subpruebas anteriores, esta puntuación se transforma en un CI, obtenido de la estandarización de Altez (1992). El presente estudio utilizó el CI.

Prueba de Habilidades Académicas (Whiteman, 1987). Para la presente investigación se adaptó el *First Grade Screening Test (FGST)*, instrumento proveniente de un proyecto estatal de intervención psicopedagógica a estudiantes de primaria de la ciudad de New York, PASS, *Program of Auxiliary Services for Students* (Whiteman, 1987). Se diseñó para ser aplicado grupalmente de una manera eficiente y rápida, y obtener así resultados confiables. El trabajo original no fue publicado para uso comercial, sino como un material interno de uso profesional en la ciudad de New York. Sobre su estructura, el instrumento contiene 46 ítems creados para evaluar las habilidades académicas de los niños ingresantes al primer grado de primaria. Los tipos de tareas incluidos en el FGST fueron de procesamiento auditivo (discriminación de sonidos iniciales y finales, rimas, comprensión auditiva, seguimiento de instrucciones), de discriminación visual (discriminación visual de figuras y palabras, reconocimiento de patrones y copia de figuras) y concepto de número (conceptos de tamaño, cantidad, numerosidad, secuencia de estímulos y diferenciación de números y letras). El tiempo de esta versión inicial tomaba cerca de 30 minutos (Whiteman, 1987). Whiteman (1987) hipotetizó una estructura tridimensional del instrumento (Procesamiento Auditivo, Discriminación Visual y Concepto de Número/Letras), pero su análisis de componentes principales (con rotación varimax) solo reveló una dimensión latente. El FGST fue adaptado ligeramente por Hirsh-Pasek, Hayson y Lescorla (1990) para usarse con niños de cinco años, al finalizar la instrucción preescolar, pero no reportaron estimaciones de confiabilidad. De acuerdo a la recomendación de Whiteman (1987), la presente investigación usó la puntuación total, cuya consistencia interna en la muestra fue $\alpha = 0.81$.

Calificaciones académicas. Estás provinieron de las calificaciones escolares registradas por las

profesoras de los niños, respecto al rendimiento escolar. Estas calificaciones o notas escolares varían entre cero (el mínimo rendimiento posible) y veinte (el máximo rendimiento posible), y se obtuvieron por medio de las evaluaciones de sus profesoras de aula, quienes usualmente crean pruebas *ad hoc* para el grupo de estudiantes. Estas evaluaciones son consideradas como pruebas formativas (Reynolds, Livingston & Willson, 2006), y muestran suficiente validez de contenido curricular (Reynolds et ál., 2006) como para ser consideradas representativas de las evaluaciones en el contexto escolar. El rendimiento escolar obtenido para el presente estudio se refiere a los primeros tres meses de instrucción escolar en primer grado, en todos los cursos del currículo en primer grado de primaria en la institución participante.

Procedimiento

Una de las primeras tareas de la investigación fue la adaptación de la Prueba de Habilidades Académicas (PHA) de Whiteman (1987). Ya que los ítems de reconocimiento de letras y palabras correspondían a las unidades idiomáticas más frecuentes del inglés, se hicieron algunas modificaciones lingüísticas que intentaron capturar simétricamente las habilidades y la estructura de la versión de Whiteman (1987). El autor del presente estudio y una colega con amplia experiencia en psicología clínica y en psicología escolar inspeccionaron el cuadernillo original para hacer las modificaciones relevantes al currículo de aprendizajes esperados para el primer grado escolar. De este modo, se decidieron modificar los estímulos e instrucciones de las tareas que dependían fuertemente del idioma. Esencialmente, se modificaron las tareas de discriminación de rimas, sonidos iniciales y sonidos finales, y de discriminación e identificación de palabras. La representación gráfica de las alternativas de respuesta, consecuentemente, también fue modificada. Finalmente, la tarea de diferenciación de números y letras fue modificada, lo que amplió las opciones de respuesta correcta para que fueran similares al número de opciones del resto de ítems del instrumento. Todas las modificaciones fueron validadas favorablemente por la opinión de algu-

nas profesoras de primer grado de la zona en la que provenían los niños.

La evaluación de los niños consistió en la aplicación de una batería de pruebas durante el proceso de admisión para iniciar el primer grado de primaria en la institución educativa participante. Esta batería consistió en las pruebas descritas en la sección Instrumentos. Las evaluaciones fueron realizadas en el horario matutino y duraron dos sesiones de veinticinco minutos cada una. Toda la batería fue aplicada grupalmente. Considerando que se ha reportado su equivalencia con la modalidad de administración individual (Buckley 1978; Koppitz, 1984; Tolor & Brannigan, 1980), el TGB-M se administró también grupalmente, entre seis a doce 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), en concordancia con las directrices para el uso apropiado de pruebas (Hambleton, 1996; International Test Commission, 2000). Dos psicólogos licenciados administraron la batería en grupos de ocho a diez niños. El proceso de evaluación se inició en noviembre y terminó en febrero, y el análisis se realizó cuatro meses después, luego de obtener las calificaciones escolares.

La evaluación de la validez incremental seguirá un enfoque de dos pasos, similar al sugerido por Hunsley y Meyer (2003). Primero, se aplicará una regresión múltiple jerárquica para evaluar el impacto del TGB-M sobre los criterios de rendimiento académico, controlando los efectos de la habilidad intelectual y las habilidades pre-académicas. Para ello, las variables de control se ingresarán juntas primero y, finalmente, el predictor de interés (TGB-M). Este proceso se repetirá para los tres criterios de rendimiento académico. Las medidas de ajuste serán R^2 y el Criterio de Información Bayesiano de Schwarz (CBIC, Schwarz, 1978), este último como una medida de ajuste basado en la reducción de la información no explicada entre los modelos comparados (Gagne & Dayton, 2002). Desde este procedimiento, se evaluará la significancia estadística del cambio en R^2 mediante la prueba F , para probar la hipótesis nula de que la diferencia es cero

en la población (Jaccard, Turrisi & Wan, 1990). Si el cambio en R^2 es estadísticamente significativo, seguidamente se evaluará su utilidad pragmática, que es cuantificada por una medida absoluta de magnitud del efecto sobre el monto de incremento, es decir, la correlación semiparcial (r_{sp} , Cohen, 1992; Hunsley & Meyer, 2003). Hunsley y Meyer (2003) sugieren que el mínimo nivel de relativo incremento en la correlación semi-parcial se halla entre 0.15 y 0.20.

Resultados

Los estadísticos descriptivos básicos y las correlaciones se encuentran en la tabla 1. Todas las correlaciones fueron estadísticamente significativas ($p < 0.05$); y las calificaciones escolares revelan una elevada varianza compartida entre ellas ($> 55\%$), lo que indica una alta dependencia aun considerando que se tratan de diferentes áreas de rendimiento. Las correlaciones entre los criterios y las variables explicativas (PHA, TGB-M y CI) muestran magnitudes similares. En el siguiente análisis se realizará la regresión múltiple jerárquica para aportar con las evidencias de validez incremental.

En todos los criterios de rendimiento académico –tanto el modelo que incluía al Bender como el modelo que no lo incluía– se rechazó la hipótesis nula de aleatoriedad de la influencia entre los predictores y el criterio. Para el modelo 1 (sin el TGB-M), los resultados con los criterios de rendimiento escolar fueron para la nota Promedio, $F(2, 103) = 21.0, p < 0.001$; para la nota en Comunicación Integral, $F(2, 103) = 17.88, p < 0.001$; y para la nota en Lógico-Matemático, $F(2, 103) = 19.69, p < 0.001$. Para el modelo 2 (en que se incluye el TGB): nota Promedio, $F(3, 102) = 20.65, p < 0.001$; nota en Comunicación Integral, $F(3, 102) = 17.46, p < 0.001$; y para la nota en Lógico-Matemático, $F(3, 102) = 17.83, p < 0.001$.

Como se deduce de la tabla 2, la inclusión del Bender como predictor produjo un cambio estadísticamente significativo en la prueba F de cada predictor, y tales cambios consistieron alrededor de un aumento del 7.5% en la varianza explicada en los criterios de rendimiento académico. El valor de la correlación semiparcial (r_{sp}) de este incremento fue

Tabla 1
Estadísticos descriptivos y matriz de correlaciones para los instrumentos administrados

	M	DE	Asim. ^a	Curt. ^b	Min.	Máx.	PHA	TGB-M	CI	Co.I	LM	CA	PS	FR	PG
Pruebas psicométricas															
Prueba de Habilidades Académicas (PHA)	17.21	5.55	-.280	-.258	3	30	1								
TGB-M	19.13	3.88	-.846	1.229	5	26	.54	1							
CI (Cattell)	101.94	12.62	.062	-.534	75	130	.45	.39	1						
Notas escolares															
Comunicación Integral (Co.I.)	15.31	3.32	-.647	-.541	7	20	.45	.49	.43	1					
Lógico-Matemático (LM)	14.76	3.10	-.600	-.295	7	20	.49	.48	.41	.83	1				
Ciencia y Ambiente (CA)	15.68	2.92	-.806	.106	8	20	.43	.45	.44	.83	.80	1			
Personal-Social (PS)	15.90	2.64	-.504	-.508	10	20	.42	.43	.40	.80	.76	.94	1		
Formación Religiosa (FR)	15.77	2.79	-.238	-.811	11	20	.46	.47	.36	.75	.75	.76	.76	1	
Promedio General (PG)	15.485	2.71	-.656	-.329	8.8	19.6	.49	.51	.45	.92	.91	.94	.92	.87	1

^a: error estándar = 0.23; ^b: error estándar = 0.46

0.28 en la Nota Promedio y Comunicación Integral, y 0.26 en Lógico-Matemático. Estos incrementos en la ecuación de regresión fueron moderadamente superiores al mínimo rango de varianza única que debería aportar un nuevo instrumento, tal como fue

sugerido por Hunsley y Meyer (2003), es decir r semiparcial entre 0.15 y 0.20.

En la tabla 2 se muestran los criterios de ajuste de los modelos examinados en cada criterio. Específicamente se muestra que el R^2_{adj} y el Criterio

Tabla 2
Estimaciones del ajuste y coeficientes beta de los modelos con y sin el SSC

Modelo	Promedio General			Comunicación Integral			Lógico - Matemático		
	CI+PHA	CI+PHA+SCC	r_{sp}	CI+PHA	CI+PHA+SCC	r_{sp}	CI+PHA	CI+PHA+SCC	r_{sp}
R^2_{adj} (R^2)	.27 (.29)	.35 (.37)	.148	.24 (.26)	.32 (.34)	.201	.26 (.27)	.34 (.32)	.112
ΔF	--	13.20**		--	12.607**		--	1.49**	
ΔR^2	--	.081		--	.082		--	.067	
CIBS	82.72	78.22		91.38	86.63		84.18	80.56	
Coeficientes									
	Beta (t)	Beta (t)	r_{sp}	Beta (t)	Beta (t)	r_{sp}	Beta (t)	Beta (t)	r_{sp}
CI	.250 (2.52*)	.180 (1.88)	.148	.273 (2.69**)	.202 (2.06*)	.201	.20 (2.04*)	.14	.112
PHA	.360 (3.63*)	.248 (2.52*)	.198	.305 (3.01**)	.194 (1.91)	.187	.38 (3.87**)	.28 (2.83*)	.221
TGB-M	--	.327 (3.63**)	.285	--	.328 (3.55**)	.281	--	.30 (3.23**)	.263

CI: Test de Inteligencia Libre de Cultura de Cattell; PHA: Prueba de Habilidades Académicas; TGB-M: Test Gestáltico Visomotor de Bender – Modificado; CIBS: Criterio de Información Bayesiano de Schwarz; t: prueba t; r_{sp} : correlación semiparcial; *: $p < .05$; **: $p < .01$

de Información Bayesiano de Schwarz (CIBS, Schwarz, 1978) convergen en indicar que la varianza explicada por el modelo que incluye al *TGB-M* aporta una mejor información predictiva del rendimiento en los criterios seleccionados. Debe observarse que las tres variables predictoras reducen efectivamente el error de predicción en el siguiente orden de impacto sobre los criterios de rendimiento: en primer lugar, Nota Promedio, seguido de la nota en Lógico Matemático, y, finalmente, en Comunicación Integral.

Los resultados sugieren que la mayor variabilidad del rendimiento escolar en las edades muestreadas es influenciada por la inclusión del *TGB-M*, cuyo impacto en las notas escolares es consistentemente elevado. En segundo lugar, la inclusión de una prueba breve de habilidades también ayuda a explicar una parte de la varianza del rendimiento escolar, más allá de la medición de las habilidades intelectuales no verbales. Finalmente, la inteligencia evaluada por una prueba de reducida carga cultural disminuye su influencia respecto a variables de contenido académico y habilidades específicas.

Discusión

Los resultados sugieren que el *TGB-M* logra incrementar la validez predictiva de las notas escolares, luego de controlar estadísticamente los efectos de la habilidad intelectual y habilidades preparatorias para el aprendizaje. Por lo tanto, la inclusión de esta medida de habilidad visomotora puede favorecer la identificación temprana de niños que probablemente tendrán problemas de rendimiento escolar, y así lograr que las mediciones en este contexto se acerquen a los estándares de exactitud que se requieren para justificar las conclusiones que se derivan de ellas (Joint Committee on Standards for Educational Evaluation, 2003). Estos resultados aportan evidencias de un solo aspecto de la validez, y se requieren otros trabajos que respalden apropiadamente las inferencias derivadas del *TGB-M* respecto, por ejemplo, a su predictividad sobre las habilidades de escritura, la identificación temprana de problemas en el aprendizaje o de problemas asociados con lesiones cerebrales.

Una figura aproximadamente exacta de las habilidades del niño antes de ingresar al colegio impacta en los beneficios consecuentes a ella, ya que favorece la provisión oportuna de servicios de apoyo (Shepard, Kagan & Wurtz, 1998). Por lo tanto, si el reducido costo de incluir al *TGB-M* conduce a este beneficio, entonces puede ser considerado rutinariamente en los procedimientos de evaluación psicopedagógica, clínica e investigación aplicada. Efectivamente, los presentes resultados dan un respaldo empírico a la utilidad del Test Gestáltico de Bender - Modificado dentro de una batería evaluativa, en condiciones que usualmente enfrenta un psicólogo escolar, es decir la evaluación de despistaje sobre las habilidades de un niño al ingresar al primer grado de educación básica. Una apropiada selección de otras variables predictoras en una batería puede complementarse con la habilidad visomotora en este proceso de evaluación. Por ejemplo, las conductas de aprendizaje (Merino, Schaefer & Worrell, 2005) han demostrado aumentar el poder predictivo de medidas cognitivas sobre el rendimiento académico estandarizado (Durbrow et ál., 2001; Merino et ál., 2005; Schaefer, Shur, Summers & McDonald, 2004).

Los presentes resultados se alinean con la literatura que desde hace décadas declara que la habilidad visomotora tiene correlatos importantes en los primeros años de escolaridad (Brannigan & Brunner, 2002; Keogh, 1969; Kulp, 1999; Norfleet, 1973; Sortor & Kulp, 2003). Especialmente, el *TGB-M* ha logrado mayores correlaciones con criterios académicos, comparado con el sistema Koppitz (Chan, 2002; Moose & Brannigan, 1997). Aunque una comparación entre ambos métodos no se ha realizado respecto a la validez incremental, se requerirán investigaciones al respecto.

Algunas consideraciones adicionales pueden agregarse al presente estudio. Las calificaciones escolares usadas en el estudio como criterios de rendimiento académico no son expresiones exactas del aprendizaje de las materias escolares, ya que otras cualidades no cognitivas también impactan en el desempeño (Merino et ál., 2005; Durbrow, Schaefer & Jimerson, 2001). También, en el contexto sociocultural del estudio, las evaluaciones hechas por profesores pueden mostrar una considerable

variabilidad en el formato y cantidad de las tareas aplicadas, aunque con una relativa homogeneidad en el contenido, pues dependen curricularmente de las normas educativas peruanas. Esto significa que las evaluaciones hechas por los profesores no están estandarizadas, y las calificaciones obtenidas combinan valoraciones cuantitativas y cualitativas de los profesores durante la asignación de las calificaciones. Estas características pueden ser más ecológicamente válidos en el contexto del estudio, pues en Perú (y como posiblemente en otros países latinoamericanos) no se tiende a aplicar rutinariamente medidas estandarizadas de rendimiento escolar.

La generalización de los presentes resultados y conclusiones deben juzgarse de acuerdo a varios aspectos. Primero, aunque se consideró que el procedimiento de evaluación fue bien monitoreado y replicable, hay variaciones particulares que un protocolo de aplicación de pruebas no controla totalmente, como las variaciones individuales de los examinados en la motivación, disponibilidad y experiencia previa con las tareas de evaluación aplicadas (Feldt & Brennan, 1989). Segundo, debe señalarse que la confiabilidad de las calificaciones escolares no fue estimada en este estudio, y se desconoce el impacto de las fuentes de error asociadas a ella. En este punto, sin embargo, la experiencia del autor en el contexto del estudio indica que las notas escolares obtenidas desde la observación de cada profesor tienden a ser consistentes, más aún si se combinan criterios cuantitativos y cualitativos cuando se asignan estas calificaciones. De este modo, un profesor puede calificar el desempeño de un niño y moderarlo con factores personales o contextuales que pueden haber disminuido o incrementado el rendimiento, pero que son irrelevantes a la materia aprendida. Aunque hay evidencias de que los profesores escolares sin cursos de medición pueden lograr mediciones consistentes (Valentin & Godfrey, 1996; Wise, Lukin & Roos, 1991), se requiere una prueba empírica actual en el contexto del estudio, dados los contradictorios informes sobre la validez y confiabilidad de las pruebas hechas por los profesores (Oescher & Kirby, 1990; Stiggins & Bridgeford, 1985; Wise, et ál., 1991).

Finalmente, el aporte predictivo de la habilidad de integración visomotora ofrece una visión útil pero parcial del papel de esta habilidad sobre el rendimiento escolar al inicio de la escolaridad. La habilidad visomotora, tal como es evaluada por el TGB-M, es una función integrativa de la habilidad motora y de la percepción visual (Beery, 2000; Brannigan & Brunner, 2002), y aunque ambas confluyen en la resolución de tareas visomotoras, no comparten elevada varianza con las puntuaciones de integración visomotora (Beery, 2000; Hudgins, 1977; Kulp, 1999; Sortor & Kulp, 2003; Volker et ál., 2010). Además de lo anterior, la validez ecológica del TGB-M debe ser evaluada, considerando que este aspecto es esencial para evaluar la utilidad de las pruebas aplicables en la práctica neuropsicológica (García-Molina, Tirapu-Uztárroz & Roig-Rovira, 2007). Por lo tanto, hay una pregunta de investigación pendiente de responder respecto a la validez incremental del TGB-M junto a medidas independientes de habilidad motora y de percepción visual.

Referencias

- Altez, I. (1992). *El test de inteligencia, Factor G, Escala I, Forma abreviada colectiva de R. B. Cattell estandarizado en Lima Metropolitana*. Lima, Perú: Universidad Femenina del Sagrado Corazón.
- Archer, R. P. & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment*, 7(3), 227-235.
- Aronholt, D. S., Frass, B. & Fraas, J. W. (1990, octubre). *Kindergarten screening test: Does it predict academic achievement?* Paper presented at the 12th Annual meeting of the mid-Western Educational Research Association, Chicago, IL.
- Beery, K. E. & Beery, N. A. (2000). *Prueba Beery-Buktenica del desarrollo de la integración visomotriz*. México, D.F.: El Manual Moderno.
- Bender, L. (1938). A visual-motor gestalt test and its clinical use. *American Orthopsychiatric Association Research Monograph*, 3.
- 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. In B. A. Bracken

- (Ed.) *Psychoeducational assessment of preschool children, 3th ed.* (pp. 33-44). Needham Heights, MA: Allyn & Bacon.
- 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. & Decker, S. L. (2003). *Bender Visual-Motor Gestalt Test, Second Edition*. Itasca, IL: Riverside Publishing.
- 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.
- Cattell, R. B. & Cattell, A. K. S. (1989). *Test de Factor "g", Escala I*. Madrid: TEA.
- 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.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Colarusso, R., Gill, S., Plankenhorn, A. & Brooks, R. (1980). Predicting first-grade achievement through formal testing of 5-year-old. *Journal of Special Education, 14*(3), 355-363.
- Coy, M. N. (1974). The Bender Visual-Motor Gestalt Test as a predictor of academic achievement. *Journal of Learning Disabilities, 7*(5), 317-319.
- deHirsch, K., Jansky, J. J. & Langford, W. S. (1966). *Predicting reading failure*. New York: Harper and Row.
- Durbrow, E. H., Schaefer, B. A. & Jimerson, S. (2001). Learning-related behaviors versus cognitive ability in predicting academic performance in Vincentian children. *British Journal of Educational Psychology, 71*, 471-483.
- Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. H. Linn (Ed.), *Educational Measurement* (3rd ed.). American Counsel of Education. New York: Macmillan.
- Foncodes (2006). *Focalización geográfica: Nuevo mapa de pobreza de Foncodes 2006*. Lima: Ministerio de la Mujer y Desarrollo Social.
- Gagne, P. & Dayton, C. M. (2002). Best regression using information criteria. *Journal of Modern Statistical Methods, 1*, 479-488.
- García-Molina, A., Tirapu-Uztárroz, J. & Roig-Rovira, T. (2007). Validez ecológica en la exploración de las funciones ejecutivas. *Anales de Psicología, 23*(2), 289-299.
- Hambleton, R. K. (1996). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. En J. Muñiz (Coord.), *Psicometría* (pp. 207-238). Madrid: Universitas.
- Haynes, S. N. & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*(4), 456-66.
- Haynes, S. N., Nelson, K. & Blaine, D. C. (1999). Psychometric issues in assessment research. In P. C. Kendall, J. N. Butcher & G. N. Hombeck (Eds.) *Handbook in research methods in clinical psychology* (pp. 125-154). New York: Wiley.
- Hirsh-Pasek, K., Hyson, M. C. & Rescorla, L. (1990). Academic environments in preschool: Do they pressure or challenge young children. *Early Education and Development, 1*(6), 401-423.
- Hudgins, A. L. (1977). Assessment of visual-motor disabilities in young children: Toward differential diagnosis. *Psychology in the Schools, 14*(3), 252-260.
- Hunsley, J. & Meyer, G. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological and statistical issues. *Psychological Assessment, 15*(4), 446-455.
- International Test Commission (ITC) (2000). *Guidelines on Test Use: Spanish Version*. Translation authorized by the Colegio Oficial de Psicólogos. ITC: Author.
- Jaccard, J., Turrissi, R. & Wan, C. K. (1990). *Interaction effects in multiple regression*. Thousand Oaks: Sage.
- Jansky, J. & deHirsch, K. (1972). *Preventing reading failure*. New York: Harper Row.
- Johnston, C. & Murray, C. (2003) Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment, 15*(4), 496-507.

- Joint Committee on Standards for Educational Evaluation (2003). *The student evaluation standards*. Thousand Oaks, CA: Corwin Press.
- Keogh, B. K. (1969). The Bender Gestalt with Children: Research Implications. *Journal of Special Education*, 3(1), 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. (1984). *El Test Gestáltico Visomotor para niños* (10ª. Ed.). Buenos Aires: Guadalupe.
- Kulp, M. T. (1999). Relationship between visual motor integration skill and academic performance in kindergarten through third grade. *Optometry and Vision Science*, 76, 159-63.
- Larsen, S. C. & Hammill, D. D. (1975). The relationship of selected visual- perceptual abilities to school learning. *Journal of Special Education*, 9(3), 281-291.
- 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.
- Mercer, C. D. (1991). *Dificultades de aprendizaje: Origen y diagnóstico*. Barcelona: CEAC.
- Merino, C., Díaz, M., Zapata, L. & Benites, L. (2006). School psychology in Peru. En S. R. Jimerson, T. O. Oakland & P. T. Farell (Eds.), *The Handbook International of School Psychology*, (pp. 299-307). Oakland: Sage
- Merino, C., Schaefer, B. A. & Worrell, F. C. (2005). Conductas de aprendizaje en el aula: La siguiente generación de evaluación e intervención. *Psicopedagogía*, 68, 148-153.
- 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.
- Norfleet, M. A. (1973). The Bender Gestalt Test as a group screening instrument for first grade reading potential. *Journal of Learning Disabilities*, 6(6), 383-388.
- Nunnally, J. C., & Bernstein, I. J. (1995). *Teoría Psicométrica* (3ra ed.). México, D.F: McGraw-Hill.
- Oescher, J. & Kirby, P. C. (1990). *Assessing teacher-made tests in secondary math and science classrooms*. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA. ERIC Document Reproduction Service No. 322169.
- Piotrowski, C. (1995). A review of the clinical and research use of the Bender-Gestalt Test. *Perceptual and Motor Skills*, 81, 1272-1274.
- Reynolds, C., Livingston, R. & Willson, V. (2006). *Measurement and assessment in education*. Boston: Allyn-Bacon/Pearson.
- Salvia, J. & Ysseldyke, J. E. (1985). *Assessment in special and remedial education* (3ª ed.). Boston: Houghton Mifflin.
- Schaefer, B. A., Shur, K. F., Summers, M. M. & MacDonald, S. L. (2004). Preschool children's learning behaviors, concept attainment, social skills, and problem behaviors: Validity evidence for Preschool Learning Behavior Scale scores. *Journal of Psychoeducational Assessment*, 22, 15-32.
- Schaefer, B. A., Shur, K. F., Summers, M. & McDonald, S. L. (2004). Preschool children's learning behaviors, concept attainment, social skills, and problem behaviors: Validity evidence for Preschool Learning Behaviors Scale scores. *Journal of Psychoeducational Assessment*, 22, 15-32.
- Schwarz, G. E. (1978). Estimating of the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153-158.
- Shepard, L., Kagan, S. L. & Wurtz, E. (1998). *Principles and Recommendations for Early Childhood Assessments*. Washington, D.C.: National Education Goals Panel.
- Sortor, J. M. & Kulp, M. T. (2003). Are the results of the Beery-Buktenica Developmental Test of Visual-Motor Integration and its subtests related to achievement test scores? *Optometry and Vision Science*, 80(11), 758-63.

- Stiggins, R. & Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22, 271-286.
- Sullivan, K., & Bowdem, S. C. (1997). Which tests do neuropsychologist use? *Journal of Clinical Psychology*, 53, 657-661.
- Telegdy, A. (1974b). The relationship between socioeconomic status and school readiness. *Psychology in the Schools*, 11, 351-356.
- Telegdy, A. (1975). The effectiveness of four reading tests as predictors of first grade achievement. *Psychology in the Schools*, 12, 4-11.
- Telegdy, G. A. (1974a). A factor analysis of four school readiness tests. *Psychology in the Schools*, 11(2), 127-133.
- Tolor, A. & Brannigan, G. G. (1980). *Research and clinical applications for the Bender-Gestalt Test*. Springfield, IL: Charles C. Thomas Publishers.
- Valentin, J. D. & Godfrey, J. R. (1996, November). *The reliability and validity of tests constructed by Seychellois teachers*. Paper presented at the Educational Research Association (Singapore) and Australian Association for Research in Education, Singapore, November 25th to 29th.
- Volker, M. A., Lopata, C., Vujnovic, R. K., Smerbeck, A. M., Toomey, J. A., Rodgers, J. D., Schiavo, A. & Thomeer, M. L. (2010). Comparison of the Bender Gestalt-II and VMI-V in samples of typical children and children with high-functioning autism spectrum disorders. *Journal of Psychoeducational Assessment*, 28(3), 187-200.
- 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.
- Whiteman, T. A. (1987). *The PASS first grade screening test: Statistical analysis and predictive validity*. Unpublished, Department of Human Development, Bryn Mawr Collegue.
- Wise, S., Lukin, L. & Roos, L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, 42, 37-42.

Fecha de recepción: 24 de septiembre de 2013
Fecha de aceptación: 18 de febrero de 2014