

## TEST GESTÁLTICO BENDER MODIFICADO Y VMI-4: COMPARACIÓN DE LA VALIDEZ INCREMENTAL

CÉSAR MERINO (\*)

Universidad de San Martín de Porres, Perú

**RESUMEN** Se evaluó la validez incremental de dos medidas de visomotricidad en relación a una medida de habilidad intelectual no verbal. Este aspecto no ha sido frecuentemente examinado en las investigaciones sobre estos constructos. Se usó el Test Gestáltico de Bender Modificado (TGBM), la Prueba de Integración Visomotora 4ta edición (VMI-4) y el DAP: IQ en 154 niños entre 4 y 8 años de dos colegios públicos. Se aplicó la regresión múltiple jerárquica de las medidas visomotoras sobre los puntajes de de la prueba de habilidad intelectual, controlando el efecto de la edad. Los resultados indican que, en comparación con el VMI-4, la habilidad visomotora medida con el TGBM contiene más varianza específica y más allá de la varianza explicada por la edad, y la varianza explicada por ambas medidas visomotoras es similar. Se discuten las implicaciones de estos resultados en la evaluación psicológica.

**PALABRAS CLAVE** sistema cualitativo de calificación, visomotor, bender gestalt test, DAP: IQ, VMI

### THE MODIFIED GESTALTIC BENDER TEST AND THE VISUAL-MOTOR INTEGRATION TEST: A TEST OF ITS INCREMENTAL VALIDITY

**ABSTRACT** The incremental validity of two metrics of the visual-motor function were tested against a metric of the non-verbal intellectual ability. This aspect has not been tested frequently in research about these constructs. The Modified Bender Gestalt Test (BGT-M) was used against the fourth generation of the Visual-Motor Integration Test (VMI-4) and the DAP:IQ test in 154, 4-8 year old students of two public schools. We used multiple hierarchical regression in the visual-motor metric over the intellectual ability test scores, controlling for age. Results show that, compared with VMI-4, the visual-motor metric with the BGT-M contains more specific variance, beyond that explained by age, while variance in the visual-motor metric was explained by both metrics alike. The implications of these results for psychological evaluations are discussed.

**KEYWORDS** psychometrics, visual-motor, Bender Gestalt Test, DAP:IQ, VMI

#### RECIBIDO CÓMO CITAR ESTE ARTÍCULO:

25 Octubre 2011 Merino, C. (2013). Test Gestáltico de Bender – Modificado y VMI: Una comparación de la validez incremental. *Psicoperspectivas*, 12(1), 183-204. Recuperado el [día] de [mes] de [año] desde <http://www.psicoperspectivas.cl>

**ACEPTADO**  
11 Julio 2012

#### \* AUTOR PARA CORRESPONDENCIA:

César A. Merino Soto. Instituto de Investigación de Psicología, Universidad de San Martín de Porres. Correo de contacto: [sikayax@yahoo.com.ar](mailto:sikayax@yahoo.com.ar).

DOI:10.5027/PSICOPERSPECTIVAS-VOL12-ISSUE1-FULLTEXT-201  
ISSN 0717-7798  
ISSNE 0718-6924

## Introducción

La importancia de la habilidad visomotora sobre el desempeño escolar y las habilidades de escritura en varias edades de nivel escolar, ha sido documentado desde hace varias décadas (Brannigan y Brunner, 2002; Koppitz, 1984; Taylor, 1999; Weil y Amundson, 1994); y durante la edad adulta, su relación con síntomas de psicopatología, como el trastorno obsesivo-compulsivo (Bloch, Sukhodolsky, Dombrowski, Panza, Craiglow, Landeros-Weisenberger, Leckman, Peterson y Schultz, en prensa) y síndrome de Tourette (Bloch, Sukhodolsky, Leckman, y Shultz, 2006) también ha mostrado suficiente varianza compartida como ayudar en la descripción de síntomas y estatus diagnóstico. La varianza compartida entre la habilidad visomotora y criterios de desempeño (clínico o escolar) ofrece soporte de validez para sus puntajes (discriminativa, o convergente), y dependiendo del objetivo del instrumento, el grado de correlación entre ellas permite estimar si los síntomas, el desempeño u otros criterios, varían de acuerdo a la variación del puntaje visomotor.

De acuerdo a los estudios de validez independientes y de aquellos obtenidos durante la construcción de las pruebas visomotoras, parecen demostrar que la habilidad para integrar las capacidades motoras y perceptuales da información relevante para tomar las decisiones diagnósticas, especialmente si se considera su poder predictivo junto a otras medidas. De todos los métodos de evaluación disponibles de la habilidad visomotora, el que parece tener más documentación empírica, popularidad y aceptabilidad en la comunidad de investigadores y profesionales por largos años, ha sido el Test Gestáltico Visomotor (TGB) (Bender 1938, 1946). Su amplio uso, sin embargo, no ha tomado en cuenta una buena racionalidad para interpretar apropiadamente sus puntajes, tal como fue tempranamente advertido por su propia autora (Bender, 1965). Esto significa que la evidencia sobre la covariación entre en TGB y criterios de interés clínico o educativo, han sido sobre-interpretados.

El TGB ha tenido modificaciones estructurales y funcionales importantes que han respondido a la creatividad de los autores para afrontar desafíos diagnósticos y descriptivos relevantes al trabajo clínico (Dana, Field y Bolton, 1983). Generalmente ambos cambios han sido interdependientes, pues las modificaciones en el número de diseños, la incorporación de un nuevo procedimiento de administración y los sistemas de calificación, han sido consecuentes uno del otro. Junto a los cambios en el sistema de calificación, la omisión de los diseños ha sido uno de las tempranas modificaciones

estructurales del TGB (por ejemplo, la modificación realizada por Keogh y Smith (1961).

Más recientemente, la inclusión de nuevos diseños para ampliar el escalamiento de sus puntajes (Brannigan y Decker, 2003) ha derivado en la nueva versión oficial del clásico TGB. Pero de las propuestas diseñadas para aplicar el TGB a niños, sin duda, el *Sistema Evolutivo de Koppitz* (1984) fue, y aún es, el método más popularizado para estimar el desempeño visomotor desde los 5 años hasta los 11 años (Brito, 2001; Lacks, 1999).

Hay nuevas versiones de la estructura y métodos de calificación del TGB que continúan presentándose, pero aparentemente aún con poco impacto en la práctica profesional y en la investigación del desarrollo infantil en el habla hispana. Esto se puede deducir de una inspección informal de la práctica profesional o del número de publicaciones que usan estas versiones. Uno de los sistemas de calificación recientes es el *Sistema de Calificación Cualitativa (SCC)* (Brannigan y Brunner, 2002), aplicable a una versión modificada del TGB. Esta versión fue modificada para niños entre 4 y 8 años, y desarrollada por Brannigan y Brunner (2002) desde trabajos derivados de De Hirsch (De Hirsch, Jansky y Kangford, 1966; Jansky y De Hirsch, 1972).

Las investigaciones que compararon la eficacia del SCC frente Sistema Evolutivo de calificación de Koppitz (1984) indicaron que los coeficientes de validez tienden a ser sistemáticamente más elevadas usando el SCC, además de una correlación positiva de ambas con la edad cronológica (Brannigan y Brunner, 2002; Parsons y Weinberg, 1993). Los resultados de las investigaciones en habla hispana con el BGT-M se han concentrado en Latinoamérica (específicamente, en Perú), y han mostrado buenas evidencias de confiabilidad por consistencia interna e inter-examinador, y de validez interna y externa con criterios relevantes al rendimiento escolar y conductas motoras (Merino, 2009, 2010, 2011a, 2011b; Merino y Benites, 2011).

La modificación al TGB (BGT-M) propuesta por Brannigan y Brunner (2002) es diferente en varios aspectos comparado con la versión de Koppitz: remueve ítems muy difíciles y poco discriminativos en niños menores de 9 años; el puntaje se interpreta hacia la calidad de la copia y no hacia el número de errores; el sistema de calificación es compatible con la idea fundamental de evaluar gestálticamente la calidad del copiado; el puntaje tiene mejor claridad en la definición del constructo de

habilidad visomotora; y la cobertura de edad es apropiada para el TGB-M, pues excluye a niños mayores de 9 años y se ocupa del rango edad sensible a los cambios rápidos en la visomotricidad.

Considerando que la edad representa un atributo biológico que impacta fuertemente en los cambios cognitivos, una de las hipótesis consistentemente corroboradas en la evaluación de la validez de constructo de las pruebas visomotoras es la relación entre la edad cronológica y sus puntajes. Bender (1965) conceptuó básicamente al TGB como una herramienta de evaluación maduracional de funciones integrativas, y por lo tanto, sensible a los cambios de edad. En las investigaciones realizadas en un amplio rango de edad, las correlaciones obtenidas han resaltado que ambas variables covarían linealmente en edades tempranas, y en magnitudes estadísticamente significativas (Bolen, 2003; Koppitz, 1984; Lacks, 1999).

Esta interpretación parece ser coherente con el estimador aplicado para ello, pues el coeficiente típicamente usando para explorar estas correlaciones ha sido el coeficiente producto de momentos de Pearson, que es una técnica habitualmente usada para establecer la validez desde métodos correlacionales (Nunnally y Bernstein, 1995). Hay más de diez posibles interpretaciones de este coeficiente (Falk y Well, 1997; Rovine y Von Eye, 1997), pero la interpretación estándar apunta a cuantificar el grado de linealidad monotonía de la relación entre dos puntajes.

Aunque la relación entre la edad y el desarrollo de la habilidad visomotora está bien documentada por la influencia maduracional del procesamiento visomotor (Bolen, 2003; Koppitz, 1984; Lacks, 1999; Cummings, Hoida, Machek, y Nelson, 2003), la variación en la estimación cuantitativa de esta relación puede ser influenciada por aspectos metodológicos o por diferencias naturales en la muestra de estudio. Por ejemplo, los diferentes instrumentos en la evaluación de la visomotricidad pueden mostrar variaciones ocasionadas por el grado de validez de sus ítems o el formato en que se presentan. Este formato de los ítems puede representarse en una forma estructurada (por ejemplo, la Prueba de Integración Visomotora, VMI-4; Beery y Beery, 2000) o una menos estructurada como el TGB. El sistema de calificación también presenta desafíos al evaluador debido a su aprendizaje, facilidad para llegar a un elevado acuerdo inter-calificadores y para integrarlo en la descripción diagnóstica del examinado (Dana et al., 1983).

Por otro lado, evolución del procesamiento visomotor está también influenciado por el impacto ambiental que acompaña al desarrollo cognitivo y social del niño (Koppitz, 1984; Lotz, Loxton y Naidoo, 2005), y ha sido bien documentado mediante las investigaciones internacionales en Europa (Aguirre, Cortadellas, y Tuset, 1988; Mazzeschi, y Lis, 2000), Medio Oriente (Rajabi, 2009; Alilo, 1993), África (Karr, 1982), y Latinoamérica (Brito y Santos, 1996; Casullo, 1988; Chang, 1990; Fernández, y Tuset, 2007). En estas investigaciones, el impacto del nivel socioeconómico sí produjo diferencias en los puntajes adicionalmente al efecto de la edad.

Koppitz (1984) sugirió que una medida de habilidad visomotora en edad escolar puede aportar en la predicción aproximada de la inteligencia, considerando que la edad impacta en ambas. La habilidad visomotora es el resultado de un proceso madurativo, y en las edades tempranas ocurre un rápido cambio que significa mejoras sustanciales en la capacidad de resolución de problemas que involucre la habilidad integrada de la percepción visual y la motricidad fina. La dimensión no verbal de la inteligencia es la más asociada conceptual y empíricamente con la habilidad visomotora (Beery y Beery, 2000; Brooks, 2009; Englund y Decker, 2011).

Dado el vínculo significativo entre la habilidad visomotora y la habilidad intelectual en la edad escolar (Knoff y Sperling, 1986; Wright y De Mers, 1982), el TGB podría ser un proxy de la capacidad global medida por la tarea visomotora del TGB en edades tempranas, tal como lo afirmó Koppitz en sus estudios (1984). Aunque en ausencia de una medida directa de la habilidad intelectual, una medida proxy basada en los puntajes del TGB parece de considerable interés, una elevada covariación entre la visomotricidad y la habilidad intelectual no verbal puede ser un resultado espurio debido a la influencia mediadora de edad, por lo cual la aplicación de métodos que controlen estadísticamente la covariación de terceras variables es altamente recomendado para evitar correlaciones ilusorias (Ferguson, 2009).

Aunque la información de validez y confiabilidad obtenida de las numerosas versiones del TGB pueden dar una idea de lo generalizable de su funcionamiento psicométrico, los cambios en las normas y/o en la estructura ponen un serio límite para establecer apropiadas equivalencias entre ellas. Las evidencias psicométricas de una versión del TGB no podrían servir como un *proxy* de otra versión modificada, y se requiere que las fuentes de validez sean aplicadas en todas ellas cuando ocurren

cambios sustanciales (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999).

La validez incremental consiste en el aumento de la información práctica y psicométrica más allá de lo aportado por los instrumentos existentes (Haynes y Lench, 2003; Haynes, Nelson, y Blaine, 1999), y que adiciona información interpretable como la contribución práctica a las ventajas que demuestra tener un nuevo instrumento. Este aspecto será examinado en la presente investigación.

Dentro del marco de los análisis correlacionales, la validez incremental tiende a cuantificarse con el uso de correlaciones semiparciales y regresión lineal jerárquica (Cohen y Cohen, 1983). El primero provee información esencialmente descriptiva y aproximada de la variabilidad compartida entre el criterio y los predictores (Carrasco et al., 2008), mientras que el método de regresión agrega la información sobre cambio en las unidades de las variables (coeficientes beta) y el cambio de  $R^2$  (Cohen y Cohen, 1983). El sentido psicométrico del análisis usando estos procedimientos es la cuantificación de la utilidad de una medida -en este estudio, las medidas de visomotricidad- que puede indicarse como una medida de validez de la misma.

Por lo tanto, el objetivo de la presente investigación es evaluar la validez incremental de dos instrumentos de evaluación de la habilidad visomotora, mediante su relación con una medida de habilidad intelectual basada en el dibujo de la figura humana.

Estas medidas de visomotricidad (TGB-M y VMI-4) convergen en el mismo objetivo de medición, pero tienen estructuras diferentes en relación a sus métodos de calificación, presentación y estructuración de estímulos. Ambas versiones no han sido comparadas respecto a la validez de constructo en una muestra hispana, y la información consecuente de esta investigación aportaría en revelar sus relaciones con la edad y la inteligencia no verbal con nuevas medidas de tales constructos.

Este estudio sobre la influencia de la habilidad visomotora en los puntajes del *Dibujo de Una Persona* (DAP: IQ, Reynolds y Hickman, 2004) es también novedoso pues el DAP: IQ es un instrumento de despistaje basado en el dibujo de la figura humana y es relativamente nuevo en el contexto hispano. Los resultados de validez del DAP: IQ con medidas de visomotricidad muestran asociaciones lineales positivas de magnitud moderada (Reynolds y Hickman, 2004); sin embargo, los estudios en población hispana apenas están emergiendo (por ejemplo, Dávila, Tello y Merino, 2009) y no

hay aún evidencias de la asociación entre ambos instrumentos, especialmente evaluados con la validez incremental.

## Método

### Participantes

Los participantes fueron 154 escolares, entre 4 y 8 años de edad, desde el nivel preescolar (4 años) hasta el 3° grado de primaria, y del turno mañana y tarde. Todos los niños presentes en el momento de la evaluación fueron seleccionados para la administración de las pruebas, tomando en cuenta la autorización de los padres y la evaluación subjetiva hecha por los evaluadores sobre de la disponibilidad conductual de los niños en la situación de evaluación.

Los niños se distribuyeron homogéneamente de acuerdo al sexo en cada grado y edad de la muestra. Proviene de dos colegios públicos de educación regular, pertenecientes a dos distritos de Lima Metropolitana. Los distritos de los colegios se caracterizan por contener familias de nivel socioeconómico medio bajo o bajo (nivel C y D; Asociación Peruana de Empresas de Investigación de Mercados, 2009), y alrededor de 60% de la población de tales distritos (Asociación Peruana de Empresas de Investigación de Mercados, 2009). En ambos colegios, las actividades curriculares son similares en contenido e intensidad, y previamente ninguno aplicó algún programa especial o *ad hoc* para la estimulación de habilidades visomotoras o cognitivas.

Tomando como referencia el *International Standard Classification of Education 97 (ISCED-97; UNESCO, 1997)*, el nivel educativo de los padres o apoderados se clasifica mayormente en los niveles de educación básica obligatoria completa (Nivel ISCED 2) y niveles post secundarios sin estudios universitarios (Nivel ISCED 3 y 4), y en mucho menor frecuencia, estudios universitarios (Nivel ISCED 5). Por medio de la información solicitada a los profesores, no se identificaron niños con discapacidad o con medicación que pudiera afectar el desempeño en las pruebas.

La similaridad de las características funcionales, estructurales y organizacionales de ambos colegios puede sugerir asumir la similitud de los niños de nuestra muestra respecto al monto y ritmo de experiencias de aprendizaje de los niños del mismo sistema escolar público en zonas urbanas. Se hizo una evaluación no estandarizada

por medio de las profesoras de los niños, sobre alguna condición especial en los niños que podría afectar su rendimiento; ellas no indicaron o no tenían conocimiento que los alumnos tenían condiciones físicas o psicológicas serias que podrían disminuir la validez del desempeño de los niños.

### Instrumento

*Test Gestáltico Visomotor – Modificado (TGB – M.* Brannigan y Brunner, 2002). Se usó una versión abreviada de seis diseños (A, 1, 2, 4, 6 y 8), aplicable a niños preescolares (4.5 años) hasta los primeros grados del nivel primario (8.5 años). Incluye un sistema de puntuación del desempeño gráfico del niño sobre las seis láminas presentadas, denominado *Sistema de Calificación Cualitativa (SCC)*. Cada lámina es calificada desde 0 (líneas aleatorias, garabateo, sin concepto de los diseños) hasta 5 (representación exacta del diseño), y usa un método de inspección global para evaluar el grado de diferenciación de los diseños reproducidos.

La investigación sobre la confiabilidad interna, test-retest e inter-jueces, y validez dan soporte a su valor métrico para la evaluación psicopedagógica (Brannigan y Brunner, 2002). Frente al Sistema Evolutivo de Calificación de Koppitz (1984), el SCC muestra correlaciones más elevadas con criterios de rendimiento escolar (Brannigan y Brunner, 2002; Chan, 2000). El manual presenta una extensa revisión de los hallazgos psicométricos, así como los criterios de calificación de cada diseño. Asumiendo un modelo equivalencia tau entre los ítems, el coeficiente  $\alpha$  para la consistencia interna de los puntajes del SCC en nuestra muestra fue  $\alpha = 0.93$ , mientras que la correlación intraclase (modelo de dos vías aleatorias) entre el autor del presente estudio y un colaborador que puntuaron los diseños reproducidos de 25 niños elegidos aleatoriamente de la muestra, fue 0.83; esta magnitud es similar o mejor al hallado en un estudio de validación anterior con el SCC (Merino, 2009; Merino y Benites, 2011).

*Prueba Beery-Buktenica del Desarrollo de la Integración Visomotriz, 4ta edición.* (VMI-4, Beery y Beery, 2000). Es un instrumento que evalúa la integración visomotora, mediante la copia de figuras geométricas ordenados evolutivamente y de dificultad creciente. Su propósito es la detección temprana de problemas en el aprendizaje, mediante la evaluación del funcionamiento visomotriz. Se presenta en una forma completa (27 diseños) para sujetos entre 3 a más años, y una forma

abreviada (18 diseños) aplicable entre 3 y 7 años. Se puede aplicar individual o grupalmente, siendo ésta última la modalidad de elección sugerida para propósitos de evaluación de despistaje.

Desde esta edición, se incluyen dos pruebas suplementarias estandarizadas para la evaluación de la percepción visual y la coordinación motriz (ambas de 24 ítems cada una), las cuales presentan los mismos diseños que la prueba de Copia, pero modificados para exigir relevantemente habilidad visual o motora, respectivamente. La prueba motora exige que el niño trace una línea sin salirse de los márgenes de los diseños, mientras que la prueba visual requiere que se elija, entre varias opciones, la figura que es igual a la del modelo.

Los ítems del VMI-4, como sus pruebas complementarias, se califican dicotómicamente (1 – 0), obteniéndose un puntaje de la suma de las respuestas correctas a los ítems. Sus normas se obtuvieron desde una muestra representativa de 2614 sujetos entre 3 y 18 años, distribuidos casi uniformemente en cada edad cronológica y empatada con el CENSO de 1990 de Estados Unidos. La confiabilidad del VMI-4 y sus pruebas suplementarias en las edades normativas es alta para la consistencia interna calculada con alfa de Cronbach (entre 0.79 y 0.89), estabilidad (tres semanas: entre 0.83 y 0.87) y acuerdo entre calificadores (entre 0.73 y 0.99).

En el proceso de validación, los ítems se seleccionaron por el procedimiento de Rasch-Wright, mostrando excelente discriminación y separación entre los sujetos e ítems en los diferentes rangos de edad (coeficientes de separación entre 0.90 y 0.99 para los ítems, y 0.75 y 0.95 para las personas). Las correlaciones entre el VMI-4 y sus pruebas suplementarias de Motricidad y Percepción Visual se mantienen entre niveles moderados y bajos (mediana: 0.33 y 0.35, respectivamente) mostrando adecuadas asociaciones divergentes entre ellas.

Respecto a los datos de validez, el VMI-4 es altamente sensible a los cambios de la edad ( $r > 0.80$ ), y sus correlaciones concurrentes con otras pruebas de visomotricidad (entre 0.29 y 0.93 con el BGT), correlaciones convergentes con logros académicos contemporáneos, y correlaciones predictivas con resultados académicos y condiciones discapacitantes futuras, han mostrado su efectividad y su valor incremental en las evaluaciones escolares. La información detallada de toda esta información se encuentra en el manual (Beery y Beery, 2000). Debido a la variabilidad de las medias

de los ítems en el grupo evaluado, en el presente estudio se aplicó la modificación Horst al KR-20 (Merino y Charter, 2010) para obtener la consistencia interna; este valor fue 0.97.

*Prueba de Dibujo de la Figura Humana para Estimación Intelectual en Niños, Adolescentes y Adultos (DAP: IQ, Reynolds y Hickman, 2004).* Esta es la más reciente prueba basada en el dibujo de una persona para estimar la habilidad intelectual aplicable en personas de 4 años hasta adultos (89 años). La instrucción para administrarlo consiste en solicitarle al examinado que se dibuje a sí mismo. Las instrucciones estandarizadas de calificación del DAP:IQ enfatizan los aspectos conceptuales del dibujo y no la calidad artística. El material consiste de un lápiz, hoja Bond A-4 y la hoja de registro. El sistema de calificación se basa en una puntuación cualitativa que varía desde 0 (parte no reconocible, incompleta o ausente) hasta 4 puntos (parte presente o cumpliendo todos los criterios anteriores); esta variación depende del ítem calificado.

Los ítems forman 23 criterios estandarizados para calificar el dibujo: siete criterios son calificados con 0 y 1; siete criterios con 0, 1 y 2; ocho criterios con niveles de 0, 1, 2 y 3; finalmente, un criterio con niveles de puntuación desde el 0 hasta el 4. El puntaje se obtiene de la suma simple de las puntuaciones alcanzadas en los criterios; luego se deriva un puntaje de CI, que es una estimación cuantitativa y confiable de la habilidad cognitiva general.

El manual reporta que los puntajes del DAP:IQ producen resultados estables ( $r > 0.80$ ) en cortos periodos de tiempo (una semana) en una muestra de 45 personas de 6 a 57 años. Recientes estudios en Latinoamérica (Dávila et al., 2009; Merino, 2009) ha confirmado que la consistencia interna es satisfactoria ( $\geq 0.79$ ), así como el acuerdo inter-calificadores e intracalificadores usando correlaciones intraclase en escolares ( $\geq 0.80$ ) y preescolares, ( $\geq 0.72$ ).

Los estudios reportados en el manual mencionan que las correlaciones de validez concurrente con el Sistema Koppitz y Goodenough-Harris fueron 0.85 y 0.86, respectivamente; y con el WISC-III,  $r_{prom} = 0.51$ ; y en el estudio latinoamericano, las correlaciones concurrentes con el Sistema Koppitz fueron elevadas,  $r_{prom} = 0.81$  (Dávila et al., 2009). En este mismo estudio, la correlaciones con una prueba de razonamiento

no verbal fueron más elevadas que con el Koppitz. La muestra normativa original del DAP: IQ fue 3090 personas de variadas regiones de los Estados Unidos.

En la presente muestra, debido a la variabilidad de las medias originado por el diferente escalamiento de los ítems (el DAP: IQ tiene ítems con variado número de opciones), la consistencia interna fue calculada con el coeficiente congénico Gilmer-Fedlt (1983), obteniéndose un valor de 0.82.

Para el presente estudio, la inclusión del DAP: IQ tuvo tres motivaciones: primero, se consideró una medida útil para hacer una estimación general del nivel de habilidad intelectual, segundo es una de las medidas actuales no verbales de inteligencia, y actualmente es parte de una investigación para su validación y estandarización en Perú.

### **Procedimiento**

La investigación fue aprobada por la Universidad de San Martín de Porres de Lima, en la cual el autor es profesor-investigador. La autorización para aplicar la investigación en la institución educativa se obtuvo de su Director luego de presentar el proyecto; los padres fueron informados y se enviaron cartas de autorización para la participación de sus hijos en los grupos seleccionados. Los participantes no recibieron compensación material por estar involucrados en la investigación.

Luego de ser aprobado por la institución escolar y por los padres, la participación de los niños, se procedió a aplicar las pruebas durante los primeros meses del periodo escolar (de marzo a mayo). Considerando que se ha reportado la equivalencia de puntajes entre las modalidades de administración individual y grupal para el TGB (Buckley 1978; Koppitz, 1984; McCarthy, 1975; Tolor y Brannigan, 1980), se administró grupalmente la versión abreviada del TGB-M, entre 6 a 8 niños por grupo. El orden de aplicación de las pruebas se modificó en cada mitad de la muestra. La batería se aplicó en una sola sesión, cuyo tiempo de desarrollo fue aproximadamente entre 16 y 25 minutos. Durante la evaluación del TGB-M, se mantuvieron las condiciones estandarizadas recomendadas para maximizar la varianza relevante al constructo evaluado (Bracken, 2000; McCallin, 2006) y en concordancia con las directrices para el uso apropiado de pruebas (Hambleton, 1996; International Test Commission, 2000).

Respecto a los análisis, se usaron los puntajes directos totales de cada prueba. Primero se analizó la varianza compartida entre las variables; esto significó ver la diferencia entre las correlaciones de las medidas visomotras con la edad y con la medida de habilidad intelectual, y para ello se aplicó una prueba para correlaciones dependientes que incluyen un elemento común (Williams, 1959). Aunque hay otras pruebas para esta situación, la  $t$  de Williams logra mantener un mejor control del Error Tipo I y más poder estadístico (Hittner, May, y Silver, 2003).

La segunda parte del análisis se hizo aplicando una estrategia de regresión múltiple jerárquica, para estimar el incremento en la correlación entre dos variables, más allá de la influencia de la otra variable (Cohen y Cohen, 1984). Se examinó el impacto de las variables visomotoras sobre la variabilidad del DAP: IQ mediante los coeficientes  $\beta$ , que permite la comparación entre variables estandarizadas. Adicionalmente, se analizaron las correlaciones semi-parciales para cada puntaje visomotor para valorar la varianza única que cada una aporta luego que se eliminara la varianza de la edad.

La evaluación de la validez incremental seguirá un enfoque de dos pasos, similar al sugerido por Hunsley y Meyer (2003). Primero, se aplicó una regresión múltiple jerárquica para evaluar el impacto del SCC y VMI-4 sobre los puntajes del DAP: IQ, controlando por los efectos de la edad. Este proceso se repitió separadamente para el SCC y VMI-4. La medida de ajuste será  $R^2$ . Desde este procedimiento, se evaluó la significancia estadística del cambio en  $R^2$  mediante la prueba  $F$ , para probar la hipótesis nula de que esta la diferencia es cero en la población (Jaccard, Turrisi, y Wan, 1990). Cuando se detectó un cambio en  $R^2$  estadísticamente significativo, se evaluó su utilidad pragmática mediante la correlación semiparcial ( $r_{spr}$  Cohen, 1992; Hunsley y Meyer, 2003), que se interpreta como el monto absoluto de incremento único de la variable de interés (pruebas visomotoras). El mínimo nivel de incremento sugerido en la correlación semi-parcial es entre 0.15 y 0.20 (Hunsley y Meyer, 2003).

## Resultados

**Análisis correlacional.** Para la variable criterio (DAP: IQ) y las variables predictivas (SCC, VMI-4 y edad), en la Tabla 1 se presentan los estadísticos descriptivos y correlacionales. Los coeficientes de confiabilidad indican un excelente control de la varianza de error, por lo tanto las correlaciones pueden alcanzar un buen poder

estadístico (Charter, 1997) al desatenuar el efecto del error de medición sobre los puntajes.

La varianza compartida entre las medidas aplicadas es mayor al 49% ( $r > 0.70$ ) de la varianza total. Las correlaciones de las variables de criterio con el SCC han tendido a ser levemente mayores que las halladas con el VMI. Aplicando una prueba  $t$  para comparar correlaciones dependientes que tienen un elemento común (Williams, 1959), la diferencia entre las correlaciones de las pruebas visomotoras y la habilidad intelectual (DAP: IQ) no fue estadísticamente significativa,  $t(151) = 1.788$ ,  $p = .07$ .

**Tabla 1**  
Información descriptiva y correlacional de los puntajes

	Mín.	Máx.	M	DE	Asd	Cue	rxxf	Correlaciones				
								DAP: IQ	SCC	VMI-4	Edad	
DAP: IQa	2	30	13.903	6.893	.293	-.952	.82	1.000				
SCCb	0	28	17.071	6.884	-.603	-.660	.93	.801	1.000			
VMI-4c	1	18	11.006	4.243	-.450	-.671	.97	.748	.801	1.000		
Edad	4	8	5.51	1.378	.533	-1.004	-	.806	.864	.804	1.000	

a: Dibujo de una Persona. b: Sistema de Calificación Cualitativa. c: Prueba de Integración Visomotora – 4 versión. d: Asimetría. e: Curtosis. f: Coeficiente de consistencia interna. Ver la sección Procedimiento para los detalles de los cálculos.

Respecto a la edad, sí se detectaron diferencias,  $t(151) = 2.457$ ,  $p = .015$ , indicando que el SCC, en comparación con el VMI-4, comparte más varianza común con la edad, y por lo tanto sus puntajes son más sensibles a los cambios de edad. Las correlaciones halladas en la Tabla 1 informan de las relaciones convergentes y divergentes entre las medidas aplicadas. La edad es similarmente asociada con todos los instrumentos, indicando que en el rango de edad muestreado, el avance de estas habilidades es fuertemente lineal y puede modelarse con el método de mínimos cuadrados. Por otro lado, las fuertes relaciones entre el DAP: IQ y las medidas de visomotricidad no

apoyan las diferencias conceptuales entre ambos constructos, pero estas correlaciones no están parcializadas por la edad.

**Análisis de regresión (validez incremental).** La Tabla 2 exhibe los resultados de la regresión jerárquica entre el DAP: IQ y las pruebas visomotoras, controlando los efectos de la edad. La edad en el primer paso de los modelos en ambas pruebas visomotoras da resultados constantes porque es la misma variable introducida. La edad muestra un elevado poder de predicción de los puntajes en el DAP: IQ,  $F(1, 152) = 267.71$ ,  $p < .001$ , de tal modo que el porcentaje de varianza explicada solamente por la edad es más del 50% de la variabilidad de los puntajes en el DAP: IQ. La edad mostró tener fuerte impacto sobre la habilidad intelectual, y la velocidad del cambio fue de cerca 4 unidades del puntaje directo del DAP: IQ ( $B = 3.99$  en la Tabla 2).

**Tabla 2**  
Resultados de la regresión múltiple jerárquica separada para el VMI-4 y el SCC

		R (R2c)	Edad			Prueba visomotora <sup>a</sup>		
			B (E.E)	$\beta$ .	$r_c$	B (E.E)	$\beta$ .	$r_c$
<b>VMI -4</b>								
Paso	1:	.79 (.63)	3.995** (.24)	.799	-	-	-	-
Edad								
Paso 2:		.81 (.66)	2.960** (.35)	.592	.386	.444** (.11)	.273	.178
Edad + VMI-4								
<b>SCC</b>								
Paso 1: Edad		.79 (.63)	3.995** (.24)	.799	-	-	-	-
Paso 2:		.82 (.68)	2.54** (.38)	.509	.305	.36** (.076)	.36	.217
Edad + SCC								

\*\* $p < 0.01$ . <sup>a</sup>:Prueba visomotora: bloque en que se ingresan los puntajes del VMI-4 y del SCC (BGT-M). R (R2c): Correlación múltiple cuadrado no corregida y corregida.  $r_{sp}$ : Correlación semi-parcial. VMI-4: Prueba del desarrollo de la integración visomotora. SCC: Sistema Cualitativo de Calificación. E.E: Error estándar.

Luego de introducir las pruebas visomotoras, el cambio producido por la edad fue de alrededor de 2.7 unidades del puntaje directo en el DAP: IQ. Como se observa en los resultados del paso 2, el  $R^2_c$  en ambas pruebas visomotoras indica una similar magnitud en la explicación de la variabilidad de los puntajes; es decir, mayor al 65% de la varianza explicada.

Considerando el VMI-4, su introducción produjo un cambio estadísticamente significativo en  $R^2_c$  ( $F[1, 151] = 14.54, p < .001$ ), adicionando 3.2% de varianza explicada única ( $r_{sp} = 0.178$ ) luego de controlar los efectos de la edad; mientras, el SCC también produjo un cambio estadísticamente significativo algo mayor en el  $R^2_c$  ( $F[1, 151] = 22.63, p < .001$ ), correspondiendo a 4.7% de varianza única ( $r_{sp} = .217$ ).

Estos resultados conciernen a la validez incremental de ambos métodos de evaluación de la visomotricidad, y muestran que el SCC permite introducir más varianza única que el VMI-4.

Finalmente, el VMI-4 y la edad muestran un aumento en el ajuste predictivo del DAP: IQ ( $F[2, 151] = 153.06, p < .001$ ); el SCC y la edad también tuvieron este efecto, pero ligeramente mejor ( $F[2, 151] = 164.23, p < .001$ ) en comparación con el VMI-4.

### Discusión

El objetivo del presente estudio fue comparar la validez incremental de dos medidas de habilidad visomotora en niños escolares. Los resultados indican que, comparado con el VMI-4, el SCC aporta más varianza única sobre los puntajes de la habilidad intelectual cuando la edad se controla estadísticamente. Esto supone que la habilidad visomotora medida por esta nueva modalidad de evaluación del TGB añade información que es útil para describir las habilidades adicionales de los niños, específicamente con el desempeño de la tarea requerida por el DAP: IQ.

La correlación semi-parcial reportada para SCC parece sugerir una mejor predicción de los cambios en los puntajes directos del DAP: IQ en el rango de edad investigado, aportando con un monto mayor de varianza única. Esto es una ventaja para el SCC. Sin embargo no resta validez al VMI-4; más bien revela que en la evaluación de las capacidades cognitivas de los niños, los puntajes derivados del SCC aportan con información específica que puede ayudar a interpretar los resultados de la evaluación clínica.

Por otro lado, la menor varianza única del VMI respecto al SCC es un resultado sorprendente, si se toma en cuenta la construcción del VMI, en la que ha representado un esfuerzo asociarlo estrechamente con los cambios debido a la edad (Beery y Beery, 2000). En el manual, para todos sus ítems hay un descriptor evolutivo que complementa la interpretación del desempeño, y que indica el nivel de edad en que aproximadamente se logra realizar el diseño. Esto hace que los ítems estén ordenados y presentados de acuerdo a una dificultad evolutivamente determinada. Si esto se mantiene consistentemente en todos los ítems, entonces el puntaje total debe estar fuertemente asociado con la edad. Los presentes resultados, sin embargo, no van en esta dirección.

Los resultados obtenidos están en la línea de las investigaciones sobre el SCC, pues este tiende a mostrar correlaciones más elevadas con medidas de rendimiento académico y de inteligencia cuando es comparado con el sistema Koppitz (Brannigan y Brunner, 2002; Chan, 2000; Parsons y Weinberg, 1993).

Un hallazgo secundario es respecto al DAP: IQ. El DAP: IQ requiere el dibujo de uno mismo para obtener un único puntaje para estimar la habilidad intelectual. Los resultados muestran que los puntajes de esta medida están fuertemente afectados por la edad. La magnitud de la relación hallada entre la edad y el DAP: IQ aporta en la validez de constructo de este nuevo instrumento, confirmando en una muestra independiente la interrelación de ambas.

Pero en esta asociación, la variabilidad debida a la habilidad visomotora también está involucrada. El 4.7% de la variabilidad de los puntajes en el DAP: IQ puede ser explicado por la habilidad visomotora medida por el SCC, y el 3.1% por los puntajes del VMI-4.

Esto sugiere también que en la calificación del DAP: IQ se debe diferenciar cuidadosamente entre ambas cualidades al asignar el puntaje a cada parte puntuable de la prueba. El sistema de calificación del DAP: IQ es único, y hace énfasis a la calidad conceptual de las partes dibujadas, independientemente de la calidad artística del producto. Debido a la que la magnitud de varianza única de ambas medidas visomotoras es pequeña, aún es considerada significativa (Hunsley y Meyer, 2003)

Hay algunos límites de la investigación que se describen a continuación. En primer lugar, la generalización de los resultados está comprometida, pues la muestra no ha

sido elegida mediante algún proceso aleatorio. Aunque las características del proceso de aplicación de los instrumentos pueden suponer la validez ecológica de los resultados para la práctica profesional, los resultados pueden ser únicos para la muestra de participantes. Se requeriría una replicación del estudio para poder dar una mayor confianza en la generalización de los resultados. En segundo lugar, no se controló eficientemente las posibles variaciones individuales en el estatus socioeconómico de cada participante; considerando que hay evidencias que esta característica covaría con los puntajes de visomotricidad y habilidad intelectual (Koppitz, 1984; Lotz et al., 2005), se requeriría una descripción empírica más completa de esta variable. También debe considerarse que los hallazgos obtenidos responden parcialmente a los aspectos únicos de la validez incremental recomendados por Haynes y Lench (2003); esto es que, primero, el TGB-M parece ayuda a discriminar el rendimiento visomotor más allá de la información de la edad, y en relación con la habilidad intelectual no verbal; y segundo, los costos de su implementación pueden ser más baratos en materiales y entrenamiento que el VMI. Pero la eficacia predictiva junto a otras medidas diferentes no ha sido investigada, y ello sugiere otras preguntas de investigación. Por ejemplo, si la habilidad visomotora evaluada por el TGB-M contribuye a detectar problemas en el aprendizaje de la escritura, lectura y/o matemáticas, y si esta capacidad predictiva es moderada por el nivel socioeconómico de los niños.

Considerando los resultados obtenidos, la versión del TGB usada en el presente estudio ofrece varias ventajas comparados con la versión que aún parece ser frecuentemente usada en el contexto hispano, es decir, la versión de 9 láminas (Bender, 1938, 1946) y el sistema Evolutivo de calificación de Koppitz (1984). Debido a varias limitaciones conceptuales (Bender, 1965) y psicométricas del TGB con el sistema de Koppitz (Knoff y Sperling, 1986), y que el TGB-M está mostrando evidencias que acumulan satisfactorias propiedades psicométricas en población hispana (ver, Merino, 2009, 2010; Merino y Benites, 2011), los presentes resultados dan soporte para que el TGB-M pueda ser considerado como una herramienta importante en una batería de pruebas psicológicas para niños de primaria.

Finalmente, ya que las pruebas usadas en el presente estudio no han sido ampliamente usadas en estudios hispanos, los resultados obtenidos también dan aportes a la validez de las medidas usadas.

## Referencias

- Alilo, M. (1993). Standardization of the Bender-Gestalt Test on 7 to 10 year-old school children in Tabriz. *Psychological Research*, 2, 33-45.
- Aguirre, G., Cortadellas, M. y Tuset, A. (1988). *Baremación del test de Bender*. Barcelona: Oikos-Tau.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.
- Asociación Peruana de Empresas de Investigación de Mercados (2009). *Niveles socioeconómicos 2009*. En línea: [http://www.apeim.com.pe/images/APEIMNSE2008\\_2009.pdf](http://www.apeim.com.pe/images/APEIMNSE2008_2009.pdf)
- Beery, K., y Beery, N. (2000). *Prueba Beery-Buktenica del desarrollo de la integración visomotriz (4ta ed.)*. México, D.F.: El Manual Moderno.
- Bender, L. (1938). A visual-motor gestalt test and its clinical use. *Research Monographs, No. 3*. New York: American Orthopsychiatric Association.
- Bender, L. (1946). *Instructions for the use of the Visual Motor Gestalt Test*. New York: American Orthopsychiatric Association.
- Bender, L. (1965). On the proper use of the Bender Gestalt Test. *Perceptual and Motor Skills*, 20, 189-190.
- Bloch, M., Craiglow, B., Dombrowski, P., Landeros-Weisenberger, Leckman, J., A., D., Panza, K., Peterson, B., Sukhodolsky, y Schultz, R. (en prensa). Poor fine-motor and visuospatial skills predict persistence of pediatric-onset obsessive-compulsive disorder into adulthood. *Journal of Child Psychology, and Psychiatry*.
- Bloch, M., Sukhodolsky, D., Leckman, J., y Schultz, R. (2006). Fine-motor skill deficits in childhood predict adulthood tic severity and global psychosocial functioning in Tourette's syndrome. *Journal of Child Psychology, and Psychiatry*, 47(6), 551-559.
- Bolen, L. (2003). Constructing local age norms based on ability for the Bender-Gestalt Test. *Perceptual and Motor Skills*, 97, 467-476.
- Bracken, B. (2000). Maximizing content-relevant variance: The assessment situation. En B. A. Bracken (Ed.) *Psychoeducational assessment of preschool children, 3th. ed.* (pp. 33-44). Needham Heights, MA: Allyn y Bacon.
- Brannigan, G. y Brunner, N. (2002). *Guide to the Qualitative Scoring System for the modified version of the Bender-Gestalt Test*. Springfield, IL: Charles C. Thomas.

- Brannigan, G., y Decker, S. (2003). *Bender Visual-Motor Gestalt Test* (2nd ed.). Itasca, IL: Riverside Publishing.
- Brito, G. y Santos, T. (1996). The Bender Gestalt Test for 5- to 15-year old Brazilian children: Norms and validity. *Brazilian Journal of Biological Research*, 29(11), 1513-1518.
- Brito, G. (2001). Bender Gestalt. In W. E. Craighead, y C. B. Nemeroff (Eds.) *The Corsini Encyclopedia of Psychology and Behavioral Science*, 3th edition (Vol. 1, pp. 201-203). New York: Wiley and Sons.
- Brooks, J. (2009). Structural extension of the Cattell-Horn-Carroll Cross-Battery approach to include measures of visual-motor integration. *Dissertations Paper 38*. Disponible en: [http://digitalarchive.gsu.edu/cps\\_diss/38](http://digitalarchive.gsu.edu/cps_diss/38).
- 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.
- Casullo, M. (1988). *Test de Bender Infantil. Normas Regionales Argentinas*. Buenos Aires: Guadalupe.
- Chan, P. (2000). Comparison of visual motor development in Hong Kong and USA assessed on the Qualitative Scoring System for the Modified Bender Gestalt Test. *Psychology Reports*, 88, 236-240.
- Chang, G. (1990). *Nueva escala de maduración del Bender Infantil*. Lima: Biblioteca Andina de Psicología.
- Charter, R. (1997). Effect of measurement error on test of statistical significance. *Journal of Clinical and Experimental Neuropsychology*, 19, 458-462.
- Cohen, J. y Cohen, P. (1984). *Applied multiple regression/correlation analysis for the behavioral sciences (2da ed.)*. Hillsdales, N. J.: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cohen, J. y Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioural sciences*. New York: Erlbaum.
- Cummings, J., Hoida, J., Machek G., y Nelson, J. (2003). Visual-motor assessment of children. In C. R. Reynolds y R. W. Kamphaus (Eds.) *Handbook of psychological and educational assessment of children: intelligence, aptitude, and achievement (2. Ed., pp. 611-653)*. New York: Guilford Press.
- Dana, R., Field, K., & Bolton, B. (1983). Variations of the Bender-Gestalt Test: Implications for training and practice. *Journal of Personality*, 47(1), 76-84.
- Dávila, M., Tello, K., y Merino, C. (2009). Nueva versión del dibujo de una persona para la estimación intelectual: Estudio psicométrico preliminar en preescolares. *Fractal: Revista de Psicología*, 21(2), 443-444.

- De Hirsch, K., Jansky, J., y Langford, W. (1966). *Predicting reading failure*. New York: Harper and Row.
- Englund, J. y Decker, S. (2011, February). *Cognitive and developmental influences in visual-motor integration skills*. Poster presented in the 2011 NASP Annual Convention, February, 22-25, San Francisco, CA.
- Falk, R., y Well, A. (1997). Many faces of the correlation coefficient. *Journal of Statistics Education*, 5(3). Disponible en línea: <http://www.amstat.org/publications/jse/v5n3/falk.html>.
- Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538.
- Fernández, T., y Tuset, A. M. (2007). Bender performance and socioeconomic status in Mexican children: A cross-cultural study. *Perceptual and Motor Skills*, 105(3), 906-914.
- Gilmer, J., y Feldt, L. (1983). Reliability estimation for a test with parts of unknown lengths. *Psychometrika*, 48, 99-111.
- Hambleton, R. (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., y Lench, H. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15(4), 456-66.
- Haynes, S., Nelson, K., y Blaine, D. (1999). Psychometric issues in assessment research. In P. C. Kendall, J. N. Butcher, and G. N. Holmbeck, *Handbook of research methods in clinical psychology* (pp. 125-154). New York: Wiley
- Hittner, J., May, K., y Silver, N. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, 130(2), 149-168.
- Hunsley, J., y 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., Turrisi, R., y Wan, C. (1990). *Interaction effects in multiple regression*. Thousand Oaks: Sage.
- Jansky, J., y De Hirsch, K. (1972). *Preventing reading failure*. New York: Harper Row.
- Karr, S. (1982). Bender Gestalt performance of Sierra Leone West African children from four subcultures. *Perceptual and Motor Skills*, 55, 123-127.

- Keogh, B., y Smith, C. (1961). Group techniques and proposed scoring system for the Bender-Gestalt Test with children. *Journal of Clinical Psychology, 17*, 172-175.
- Knoff, H. y Sperling, B. (1986). Gifted children and visual-motor development: A comparison of Bender-Gestalt and VMI test performance. *Psychology in the Schools, 23*, 247-251.
- Koppitz, E. (1984). *El Test Gestáltico Visomotor para niños* (10ma.Ed.). Buenos Aires: Guadalupe.
- Lacks, P. (1999). *Bender Gestalt screening for brain dysfunction*, (2d ed.), New York: John Wiley y Sons,
- Lotz, L., Loxton, H., y Naidoo, A. (2005). Visual-motor integration functioning in a South African middle childhood sample. *Journal of Child y Adolescent Mental Health, 17(2)*, 63-67.
- Mazzeschi, C., y Lis, A. (2000). The Bender-Gestalt test in an Italian sample: An analysis of Koppitz's developmental Bender scoring System deviations. *Perceptual and Motor Skills, 90*, 373-385.
- McCallin, R. C. (2006). Test Administration. In S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 625-652). Mahwah, NJ: Lawrence Erlbaum.
- McCarthy, D. (1975). The feasibility of a group Bender-Gestalt test for preschool and primary school-aged children. *Journal of School Psychology, 13(2)*, 134-141.
- Merino, C. y Charter, R. (2010). Modificación Horst al coeficiente KR-20 por dispersión de la dificultad de los ítems. *Revista Interamericana de Psicología, 44(2)*, 274-278.
- 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. (2011a). Validez comparativa de tres sistemas de calificación para el Test Gestáltico Vismotor de Bender Modificado. *Revista de Psicología – UCV, 13(1)*, 90-102.
- Merino, C. (2011b). Exploración de diferencias normativas en el Sistema de Calificación Cualitativa para el Test Gestáltico de Bender Modificado. *Liberabit, 17(2)*, 199-209.
- Merino, C., y 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.
- Nunnally, J., y Bernstein, I. (1995). *Teoría psicométrica (3ra. ed.)*. México, D. F.: McGraw-Hill.

- Parsons, L. y Weinberg, S. (1993). The Sugar Scoring System for the Bender Gestalt Test: An objective approach that reflects clinical judgment. *Perceptual and Motor Skills*, 77, 883-893.
- Rajabi, G. (2009). Normalizing the Bender Visual-Motor Gestalt Test among 6-10 year-old children. *Journal of Applied Sciences*, 9(6), 1165-1169.
- Reynolds, C. y Hickman, J. (2004). Draw-a-person intellectual ability test for children, adolescents and adults (DAP-IQ). Austin: PRO-ED.
- Rovine, M. y Von Eye, A. (1997). A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. *The American Statistician*, 51, 42-48.
- Taylor, K. (1999). Relationship between visual motor integration skill and academic performance in kindergarten through third grade. *Optometry and Vision Science*, 76 (3), 159-163
- Tolor, A. y Brannigan, G. (1980). *Research and clinical applications for the Bender-Gestalt Test*. Springfield, IL: Charles C. Thomas Publishers.
- UNESCO (1997). International Standard Classification of Education (ISCED). Consultado el 1 de enero, 2011. Disponible en: <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>
- Weil, M., y Amundson, S. (1994). Relationship between visuomotor and handwriting skills of children in kindergarten. *American Journal of Occupational Therapy*, 48 (11), 982-988.
- Williams, E. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society*, 21, 396-399.
- Wright, D. y DeMers, S. (1982). Comparison of the relationship between two measures of visual-motor coordination and academic achievement. *Psychology in Schools*, 19, 473-477.