

INVESTIGACIONES

Validez de Constructo de una Prueba Breve de Pre-escritura en Preescolares

Construct Validity of a Short Test of Pre-writing to Preschoolers

César Merino-Soto^a, Gustavo Calderón-De la Cruz^b

^a Universidad de San Martín de Porres
Correo electrónico: cmerinos@usmp.pe

^b Universidad de San Martín de Porres
Correo electrónico: gcalderond@usmp.pe

RESUMEN

El objetivo de la presente investigación es evaluar psicométricamente las propiedades de un instrumento para describir el nivel de habilidad pre-escritora, PE-56 (Condemarin & Chadwick, 1991). La muestra fue conformada por 77 preescolares de alrededor de 5 años de edad. Se hizo una modificación de su sistema de calificación, mediante una adaptación del Sistema Global de Calificación del Bender-II. Se aplicó el PE-56 en grupos pequeños, y se analizó la estructura interna, confiabilidad y relaciones con criterios externos, para lo cual se aplicó el Bender-II y VMI-4. Se halló buena confiabilidad (0.82) y un modelo unidimensional subyacente a los ítems, pero con covariación de errores entre dos ítems; la correlación con pruebas visomotoras fue diferente de acuerdo al método de calificación, obteniendo distintas correlaciones entre el rendimiento en la prueba de pre-escritura y las pruebas visomotoras. Se discute el uso y límites del instrumento.

Palabras claves: pre-escritura, validación, visomotricidad, Bender Gestalt test, evaluación.

ABSTRACT

The objective of this research is to evaluate psychometrically the properties of an instrument to describe the pre-writing skill level, PE-56 (Condemarin & Chadwick, 1991). The sample was composed of 77 preschool children of about 5 years old. The Scoring System was modified by an adaptation of the Global Scoring System Bender-II. PE-56 was applied in small groups, and internal structure, reliability and relationships with external criteria (Bender-II and VMI-4) were analyzed. Good reliability (0.82) and an underlying unidimensional model of the items were found, but with error covariance between two items; the correlation with visual-motor tests was different according to the scoring method in these measures, obtaining different correlations between performance of pre-writing test and visual-motor tests. The use and limits of the instrument are discussed.

Keywords: pre-writing, validation, visual-motor function, Bender Gestalt Test, assessment.

1. INTRODUCCIÓN

La escritura es una de las herramientas de comunicación de mayor impacto en todas las sociedades del mundo. La importancia de su estudio radica en la predicción de los aprendizajes de orden superior como las matemáticas, la lectura, el dictado, la comprensión y la ortografía (Chu, 1997) y, consecuentemente, con el logro escolar (Cahill, 2009; Volman, Schendel & Jongmans, 2006). En ese sentido, la escritura está vinculada al proceso de adaptación del escolar y se convierte en reto para los profesionales de la educación, quienes orientan sus esfuerzos en estimular las habilidades maduracionales (p. ej., motricidad fina) y cognitivas (p. ej., atención sostenida; Case-Smith, Holland & Bishop, 2011) que están asociadas a su desarrollo, y en evaluar periódicamente sus niveles para detectar posibles déficits (Vinter & Chartrel, 2010). No obstante, los niveles de prevalencia del déficit escritor en alumnos son elevados y oscilan entre un 10% y 30% (Feder & Majnemer, 2007; Vinter & Chartrel, 2010), siendo los varones de mayor incidencia (Volman et al., 2006), y esto probablemente por la tardía detección del problema, más aún cuando es común la falta de atención de las etapas previas de la escritura (e. g., preescritura) (Tompkins, 2001 Cit. por Hashempour, Rostampour & Behjat, 2015) durante el nivel preescolar. Ante este panorama, incorporar instrumentos de medición direccionados al nivel preescolar que sean sensibles a la detección temprana de posibles déficits de la escritura y que cuenten con estándares psicométricos que certifiquen su utilidad es una alternativa de prioridad.

Al ser la escritura una habilidad compleja (Case-Smith et al., 2011; Weil & Cunningham, 1994), su abordaje requiere de minuciosidad y exhaustividad. Comúnmente sus formas de evaluación se han centrado en el rendimiento del alumno considerando la fase evolutiva (Condemarin & Chadwick, 1991; Feder & Majnemer, 2007) y la composición escrita, esta última diferenciada según el proceso (p. ej.: la planificación) y el producto (p. ej.: legibilidad del escrito). Justamente, en la evaluación de producto, sus formas de medición denominadas a) método holístico, que valora la calidad producto escrito a través del juicio global del calificador tomando de guía diseños estandarizados que cursan de lo legible a lo ilegible y b) el método de análisis discreto, que centra su examinación en los detalles del producto escrito, por ejemplo, espacios entre las letras y palabras, la formación de las letras, rectitud de la línea escrita (Rosenblum & Livneh-Zirinski, 2008; Rosenblum, Weiss & Parush, 2004), pueden ser integradas, en su estrategia de evaluación, los componentes que explican la escritura (como las habilidades cognitivas, kinestésicas o de habilidad perceptomotoras), más aún cuando se ha reportado que esto no es frecuente en los instrumentos de medición de la escritura y es necesaria su inclusión (Rosenblum, Weiss & Parush, 2003; Rosenblum & Livneh-Zirinski, 2008), dado que aportaría a una detección temprana que permita precisar con mayor agudeza las causas de un pobre desempeño escritor.

Al inspeccionar los catálogos de las editoriales hispanas sobre la evaluación de pruebas psicológicas estandarizadas únicas para la escritura, se puede hallar que este tipo de instrumento es menos frecuente que los diseñados para la lectura o matemáticas en niños y adultos. Existen algunas pruebas editadas (Cueto, Ramos & Ruano, 2004; Toro & Cervera, 2002; Urío, Toro & Cervera, 2000) que contienen subtests que evalúan directamente la calidad de escritura, pero actualmente no se han revalidado sus propiedades psicométricas ni actualizado sus normas. También, son instrumentos cuya aplicación es moderadamente prolongada, y la calificación requiere un monto de tiempo y entrenamiento

que puede desanimar a los profesionales de la educación. En general, aquellas pruebas que evalúan el producto escrito pueden requerir de mayor cantidad de tiempo y entrenamiento sostenido y guiado, a fin de familiarizarse con el procedimiento de calificación y alcanzar niveles elevados de confiabilidad inter-calificador. Como respuesta a esta limitación, los profesionales pueden involucrarse en el aprendizaje informal o casual, y recurrir a la autoinstrucción cuando utilizan una nueva prueba de evaluación (Merino, Calderón-De la Cruz & Manzanares, 2016). Una alternativa ante este panorama es diseñar instrumentos de medición efectivos en la exploración de la legibilidad de producto escrito, que sean breves para su aplicación y calificación, minimizando la rigurosidad del entrenamiento formal, pero con significancia objetiva en el proceder del puntaje a seleccionar.

Justamente, una de las propuestas de evaluación es la *Prueba de Escritura para Niños de 5 y 6 años* (PE-56, Condemarín & Chadwick, 1991). Las autoras la presentan como una prueba informal no estandarizada, para registrar la habilidad de escritura del niño mediante el copiado de letras cursivas del alfabeto español. El PE-56 fue diseñado como un método complementario al proceso de caracterización de las habilidades que un niño básicamente debe desarrollar antes de iniciar el entrenamiento de la escritura en los primeros años de la escolaridad; sin embargo, tiene componentes estandarizados como el repaso, el completamiento y copiado de letras y el copiado de una oración, que están destinados a evaluar la calidad de la escritura. Si bien no fue creado para señalar dificultades específicas, su utilidad está orientada hacia la descripción de la habilidad pre-escritora del niño, y claramente sirve como una prueba potencial de despistaje de problemas de escritura durante la etapa preescolar. Por lo tanto, el desempeño del niño en el PE-56 sería un indicador importante de déficit en la habilidad motora fina. Asimismo, su brevedad lo hace elegible para ser incluida dentro de una batería de instrumentos que evalúen las habilidades preparatorias de los niños al ingresar al primer grado de primaria, siendo una medida relevante que contribuye para evaluar el progreso del estudiante.

En la presentación del instrumento (Condemarín & Chadwick, 1991), las autoras no reportaron estudios de validez y confiabilidad en su publicación original, desconociéndose sus características métricas en la población de baremación, y limitando la evaluación de su generalización a otros grupos de niños de la población general. Como herramienta informal, tal como la clasifican las autoras (Condemarín & Chadwick, 1991), es plausible pensar que originalmente la PE-56 fue creada inductivamente, utilizando como fuentes las prácticas de enseñanza de escritura realizadas antes del ingreso al primer grado, y las reconocidas competencias de las autoras. Dado el objetivo del instrumento, parece justificable que satisfaga la validez de contenido, aspecto crítico de las evaluaciones hechas por profesores para sus clases (Fives & DiDonato-Barnes, 2013; Nitko, 2001), quienes comúnmente hacen uso de herramientas informales de evaluación (Feder & Majnemer, 2007; Nitko, 2001). La validez de contenido permite asegurar la representatividad, comprensión y la relevancia de las unidades o ítems (Mokkink et al., 2010), pero no son suficientes para hacer una interpretación válida de los puntajes cuando se quiere plantear un nuevo uso al instrumento. Otros aspectos de la validez son necesarios para asegurar la relación del instrumento con el constructo y con los usos que se pretende dar (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999).

En cumplimiento con dicho propósito, el presente estudio centró su atención en obtener evidencias de validez del PE-56, pero incluyendo un nuevo método de calificación

de sus resultados. Este método está asociado a la evaluación de la visomotricidad basada en el Test Gestáltico Visomotor de Bender (TBG; Bender, 1938), tomando en cuenta que la visomotricidad es una de las variables que comparte mayor varianza explicada dentro de los componentes de la escritura (Beery, 2000; Maldarelli et al., 2015; Rosenblum & Livneh-Zirinski, 2008; Tseng & Chow 2000; Volman et al., 2006). El nuevo modelo de calificación está influenciado por el trabajo de Brannigan & Brunner (2002), que une un enfoque holístico en la creación del *Sistema de Calificación Global* (SCG) para evaluar la calidad del diseño reproducido como producto completo y configuración total (Brannigan & Brunner, 2002; Brannigan & Decker, 2003; Merino, Allen & Decker, 2013). El método holístico y el SGC han sido implementados en las recientes versiones del TBG, específicamente en el Test de Bender Modificado (Brannigan & Brunner, 2002) y el Test Gestáltico Vismotor de Bender-2da versión (Brannigan & Decker, 2003). Este método es recomendado comparado por sobre los métodos discretos de calificación, especialmente aquellos basados en el error (e. g.: Sistema Evolutivo; Koppitz, 1984), pues presentan dificultad para capturar apropiadamente el constructo de visomotricidad (Merino & Allen, 2014), y no capturan realísticamente la continuidad del constructo, es decir, de la habilidad visomotora.

Por lo tanto, el objetivo de la presente investigación es evaluar psicométricamente las propiedades métricas del PE-56, creada para niños que están en la etapa preparatoria para el primer grado de primaria, adecuando el método holístico de evaluación visomotora del TBG basado en sus recientes versiones (Brannigan & Brunner 2002; Brannigan & Decker, 2003). Este nuevo método para calificar el desempeño del niño puede ser no solo más preciso, sino también eficiente en su implementación para la investigación y la intervención psicopedagógica.

2. MÉTODO

2.1. POBLACIÓN Y MUESTRA

Fueron 77 niños (36 mujeres, 46,8%) provenientes de una institución educativa pública, ubicada en un distrito completamente urbano, en Lima Metropolitana. La institución fue elegida debido al acceso y a la disponibilidad para aplicar los instrumentos del presente estudio, ya que se requería que cada niño resuelva las tareas en una situación de evaluación individual. Para obtener mayor poder estadístico, se muestreó a todos los niños asistentes de cada aula de la institución, correspondientes a los niveles preescolares de 4 y 5 años. El nivel socioeconómico de los niños tiende a ser medio bajo, como es frecuente en las familias que asisten a instituciones públicas de educación (Merino et al., 2006). La edad fue entre 48 y 75 meses de edad ($M = 63.1$, $DE = 6.3$), y se detectaron moderadas diferencias de edad ($t[75] = 2.61$, $p = 0.01$, $d = 0.62$) entre varones ($M = 61.3$, $DE = 6.6$) y mujeres ($M = 65.0$, $DE = 5.46$).

2.2. INSTRUMENTO

Prueba de Escritura para niños de 5 y 6 años (PE-56; Condemarin & Chadwick, 1991)

Prueba presentada como instrumento gráfico no estandarizado por Condemarin y Chadwick (1991) para evaluar la escritura preparatoria en niños entre 5 y 6 años. Los estímulos

consisten en cuatro ítems, los tres primeros contienen ocho letras (b, ch, l, m, p, r, u, y) alineadas, diseñadas en cursiva; independientemente, cada ítem requiere ser repasado, completado y copiado, respectivamente. Cada letra del primer ítem se presenta con líneas completas, debajo de las cuales el niño las reproduce; en el segundo ítem, las letras están con líneas entre-cortadas; el tercer ítem presenta las letras con líneas completas y una raya debajo de ellas para que el niño las copie sobre ellas. El cuarto y último ítem es una frase (“El niño juega en el patio”) que requiere ser copiada. El orden de los ítems parece corresponder a la dificultad para realizarlos. Aunque no se aclara en las descripciones de los autores, la aplicación es preferentemente individual, pero puede adaptarse a la aplicación grupal. Los autores no indican cómo eligieron las letras de los tres primeros ítems ni la frase que se presenta en el cuarto ítem.

Test Gestáltico Visomotor de Bender, 2da versión (Bender-II; Brannigan & Decker, 2003)

Se utilizó la nueva versión que consta de 16 láminas, aplicable a sujetos de 4 a 85 años de edad. El Bender-II tiene dos subtest principales (Copia y Recuerdo) y dos pruebas suplementarias (motricidad fina y percepción visual). La fase de Copia requiere que el sujeto reproduzca cada diseño presentado, y en la fase de Recuerdo se solicita al evaluado que recuerde y dibuje los diseños presentados. Complementariamente, se pueden aplicar subpruebas de coordinación motora y discriminación visual. Los diseños reproducidos se califican con un puntaje de 0 a 4, usando el Sistema de Calificación Global (SCG). El manual reporta buenas evidencias de validez y confiabilidad (Brannigan & Decker, 2003). Recientes investigaciones en Latinoamérica arrojan satisfactorios niveles de acuerdo intercalificadores en el nivel del puntaje y de los ítems (Merino, 2012).

Prueba Beery-Buktenica del Desarrollo de la Integración Visomotriz, 4ta edición (VMI-4; Beery, 2000)

Instrumento que evalúa la integración visomotora, mediante la copia de figuras geométricas ordenadas de creciente dificultad. Desde la 4ta 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. La información detallada de toda esta información de halla en el manual (Beery, 2000).

2.3. PROCEDIMIENTO

La naturaleza continua del desempeño motor del niño, reflejado en las respuestas a cada ítem del PE-56, fue el marco para crear una modalidad de calificación fundamentada en la calidad de la copia, cuantificado por un puntaje ordinal para cada ítem y que corresponde a un proxy de la continuidad de la habilidad pre-escritora evaluada por su calidad. Se modeló

este enfoque usando los trabajos de Brannigan y Brunner (2002) y Brannigan y Decker (2003) para las nuevas versiones del Bender Gestalt Test Modificado y la segunda versión (respectivamente). Ambos métodos usan un enfoque global para calificar la calidad de la copia de sus diseños, dando menos peso a la observación de los detalles específicos para decidir la calidad del dibujo. Para la aplicación del método de calificación global, la unidad de análisis fue cada fila de letras, obteniéndose de esta manera una puntuación global por cada fila. Se tomó esta decisión bajo varias consideraciones: facilidad en la calificación; coherencia con el objetivo de la evaluación informal de la pre-escritura y mayor confiabilidad entre calificadores; menor sensibilidad a los aprendizajes específicos e idiosincrásicos de los niños respecto a alguna letra y el uso de la prueba, es decir, descripción e identificación temprana. El sistema de calificación creado, por lo tanto, tuvo las siguientes características: estar orientado hacia la calidad de la copia, obtenerse mediante la observación global del producto, estar influenciado por el juicio y experiencia del calificador, retener con buen balance eficiencia-exactitud, y ser interpretable unidimensionalmente. La puntuación se definió en un rango de 4 puntos, del 0 al 3: 0 (Reproducción irreconocible), 1 (Reproducción con severas distorsiones), 2 (Reproducción con menores distorsiones) y 3 (Reproducción casi exacta).

Por otro lado, la administración de las pruebas se hizo individualmente, efectuada por tres estudiantes de psicología, previamente entrenadas en la aplicación y calificación de los instrumentos. Se aplicaron también las pruebas de visomotricidad en un aula de la escuela con suficiente espacio y sin distracciones relevantes. El orden de presentación de los instrumentos se alternó para evitar el efecto del orden de las pruebas (Jensen, Watanabe & Richters, 1999). Previamente, se tramitaron las autorizaciones para proceder con la recolección de datos por parte del director de la institución y de los padres de familia.

El análisis examinó la estructura interna del instrumento, la confiabilidad de sus puntajes e intervalos de confianza (Domínguez & Merino, 2015), y su relación con criterios de desempeño visomotor, motor y perceptual visual. Para evaluar la estructura interna, se usó el análisis factorial confirmatorio mediante metodología de ecuaciones estructurales, en que para el ajuste del modelo a los datos se utilizó la función χ^2 con el estimador máxima verosimilitud. Para decidir sobre el modelo que representa mejor a los datos, se aplicaron varios indicadores de ajuste (TLI > 0.95; CFI > 0.95; RMSEA < 0.05; Bentler, 1990). Se aplicó el programa informático de Boomsma y Herzog (2008) para utilizar una corrección a los resultados (Swain, 1975), recomendado para situaciones en que el tamaño muestral es pequeño (Herzog & Boomsma, 2009). Todos los análisis se realizaron usando el nivel alfa de 0.05, y con el programa EQS 6.1 (Bentler & Wu, 2004).

3. RESULTADOS

3.1. ANÁLISIS PRELIMINAR

En la Tabla 1 se muestran las características estadísticas de los ítems. La dispersión de los ítems muestra ser similar, y el nivel medio de habilidad pre-escritora se encuentra en el punto 2 (reproducción con menores distorsiones) en los ítems 1 y 2, y alrededor del punto 1 (reproducción con severas distorsiones) de la escala de calificación. Hay una aparente disminución de la calidad del rendimiento entre los ítems, que es una característica esperable

debido a la complejidad creciente de los ítems. El rendimiento en los ítems, sin embargo, no siguió un patrón lineal, sino más bien una tendencia cúbica ($F[1, 76] = 84.76, p < 0.01, \eta^2 = 0.52$); es decir, dos rendimientos altos en el ítem 2 y 4 comparados con los ítems 1 y 3, respectivamente. Los coeficientes de asimetría (g_1) y curtosis (g_2) (Fisher, 1950), así como el coeficiente Jarque-Bera (1980, 1987) indican que la normalidad distribucional de los ítems es aceptable ($\chi^2 > 5.99$ en $\alpha = 0.05$). Estos resultados justificaron la obtención de correlaciones Pearson inter-ítem para el presente análisis, ya que la asimetría y curtosis univariada de los ítems estuvieron generalmente centradas en cero, y por lo tanto pueden ser suficientes indicadores de la covariación entre los ítems (Ory & Mokhtarian, 2010).

Tabla 1. Estadísticos descriptivos y correlaciones Pearson inter-ítem (n=77)

	M	DE	. g_1	. g_2	Z_{JB}	Item1	Item2	Item3	Item4
Ítem1	1.58	1.030	.026	-1.157	4.573	1.000	.706	.545	.361
Ítem2	2.00	.778	-.34	-.401	2.129		1.000	.579	.456
Ítem3	1.09	.934	.411	-.757	4.208			1.000	.659
Ítem4	1.26	.965	.533	-.594	4.988				1.000
Total	5.935	3.015	.244	-.520	1.717				

Nota: . g_1 = coeficiente de asimetría; . g_2 = coeficiente de curtosis; Z_{JB} = estadístico Jarque-Bera para la normalidad de los datos.

3.2. ESTRUCTURA INTERNA

Se hipotetizó un modelo unidimensional para explicar las relaciones entre los ítems. El modelo inicial mostró pobres coeficientes de ajuste (modelo no modificado en la Tabla 2); examinando los índices de modificación y el efecto de los cambios indicados sobre el parámetro modificado, se observó que el único cambio estadísticamente importante sería liberando la covarianza entre los errores de los ítems 3 y 4; efectivamente, la correlación entre ambos es una de las más elevadas comparada con las demás correlaciones inter-ítem (ver Tabla 1). La varianza específica que aportan estos dos ítems parece coherentemente relacionada debido a la dificultad creciente que exigen al niño, y que lo llevaría a aplicar estrategias motrices similares durante su ejecución. Con estas consideraciones estadísticas y lógicas se liberó la covariación de error entre los ítems 3 y 4, resultando en una covariación estandarizada (correlación) latente entre ambos de 0.509 ($p < 0.05$). Esta modificación produjo un inusual ajuste perfecto. Los resultados no estandarizados de los parámetros para los ítems aparecen en la Figura 1, y sus valores estandarizados (pesos de regresión) en la Tabla 3. Las cargas son elevadas ($\lambda > 0.65$; Briggs & MacCallum, 2003) e indican que cada ítem aporta bien al constructo; se observa que los ítems de repaso y completado de letras son moderadamente más elevados que los de copiado de letras y de una oración; y el copiado de una oración es el de menor carga comparado con el resto.

Tabla 2. Correlación ítem-test corregido y cargas factoriales (solución estandarizada)

		Modelo no modificado		Modelo modificado	
				Sin ajuste Swain	Ajuste Swain ^a
Ajuste					
χ^2 (g.l. = 2)	-	19.413		1.009	.981
RMSEA [I.C. 90%]	-	0.338 [.21, .48]		0.0 [.0, .18]	0.0 [.0, .18]
TLI	-	.588		1.0	1.0
CFI	-	.863		1.0	1.0
Parámetros	r_{itc}^b	λ	R^2	λ	R^2
Ítem 1	.622	.699	.327	.808	.244
Ítem 2	.708	.944	.382	.873	.445
Ítem 3	.728	.618	.890	.667	.763
Ítem 4	.569	.572	.488	.494	.653

Nota: ^a factor Swain de corrección fue 0.972. ^b correlación ítem-test corregida. λ = carga factorial estandarizada. R^2 = confiabilidad del ítem.

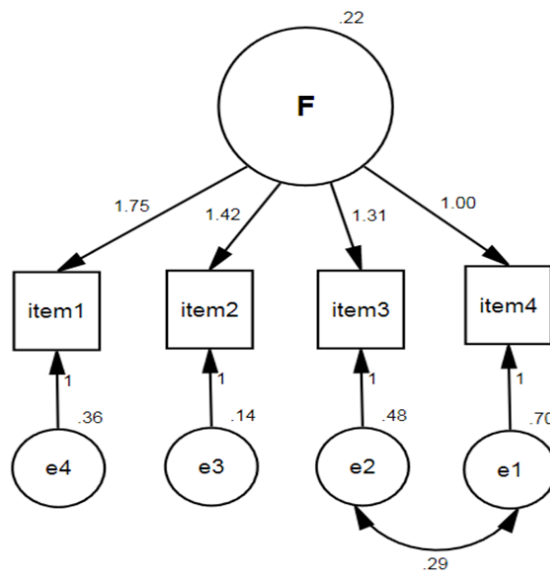


Figura 1. Parámetros no estandarizados del modelo de medición

3.3. CONFIABILIDAD

La consistencia interna por el método α (Cronbach, 1951) fue .82 (I.C. 95% = .74 – .87), lo que indica que los ítems generan respuestas consistentes en alto grado. Debe observarse que con el pequeño número de ítems y la naturaleza interrelacionada de las conductas medidas con ellas, podría esperarse aún más altos coeficientes de consistencia interna; sin embargo el coeficiente fue bajo 0.85, lo que podría sugerir un monto de varianza única de cada ítem y menor redundancia entre ellas.

3.4. VALIDEZ CON CRITERIOS EXTERNOS

Las correlaciones entre la prueba de pre-escritura y sus ítems con las medidas de percepción visual (Bender-II y VMI-4) no fueron estadísticamente significativas, y la máxima varianza compartida entre ellas fue apenas 7% (con Bender-II) y 2.4% (con VMI-4); en cambio, sus correlaciones con las medidas de visomotricidad y motricidad del Bender-II y VMI-4 fueron sistemáticamente mayores y estadísticamente significativas. También se observa que un patrón correlacional diferente: en el Bender-II, el puntaje Copia tendió a presentar más varianza compartida con los ítems (exceptuando el ítem 2, entre 8.8% y 17.22%) y puntaje total de la prueba de pre-escritura; en cambio, en el VMI-4, la mayor varianza compartida ocurrió entre su puntaje Motor y los ítems (exceptuando el ítem 4, entre 12.4 y 24.4%) y puntaje total de la prueba de pre-escritura.

Tabla 3. Correlación entre la prueba de escritura y medidas de visomotricidad (Bender-II y VMI-4)

	Bender-II			VMI-4		
	Copia	Visual	Motor	Copia	Visual	Motor
Ítem 1	.298**	.107	.259	.247*	.121	.456**
Ítem 2	.372**	.218	.418**	.309**	.072	.494**
Ítem 3	.481**	.278	.434**	.322**	.137	.353**
Ítem 4	.415**	.277	.324*	.327**	.155	.216
Total	.480**	.268	.436**	.369	.153	.462**

Nota: *p < .05. **p < .01

4. DISCUSIÓN

El presente estudio reporta la validación de una propuesta de calificación para la prueba de pre-escritura elaborada por Condemarin y Chadwick (1991). Esta propuesta de calificación fue modificada para hacerla congruente con la continuidad del constructo del desempeño motor, basado en los recientes trabajos sobre las modificaciones realizadas en del Test Gestáltico Visomotor de Bender (Brannigan & Brunner, 2002; Brannigan & Decker, 2003).

Se halló que las tareas de copia, trazado y repaso involucradas en la prueba de pre-escritura se influyen por la capacidad motora fina del niño, así como por la capacidad visomotora; y, en contraste, aparece menos influenciada independientemente por la percepción visual.

También se halló que la covariación entre las pruebas visomotoras fue diferente de acuerdo a su método de calificación, obteniendo diferentes correlaciones entre el rendimiento en la prueba de pre-escritura y el Bender-II comparado con el VMI-4. Parece que el procedimiento de calificación continua adaptado a la prueba de pre-escritura permite una mayor covariación no solo con este constructo, sino también con otros relevantes, un aspecto reconocido en otros estudios que comparan los métodos de calificación global y métodos discretos en el TGB (Chan, 2002; Keogh & Smith, 1971; Merino, 2011, 2013). Esto tiene sentido si se toma en cuenta el efecto de la varianza del método, una consecuencia que puede producir los diferentes métodos de recolección de datos sobre la variabilidad explicada.

Otro resultado diferencial fue la tendencia a una mayor relación entre la medida motora del VMI-4 comparado con el subtest motor del Bender-II. Esto puede explicarse de acuerdo a dos puntos: primero, la subprueba motora del Bender-II generalmente presenta el efecto de techo, pues la mayoría de los participantes obtiene puntajes altos, restringiéndose enormemente su varianza; pero esto es razonable porque el subtest motor fue creado como medida de despistaje y sus ítems fueron seleccionados para ser fácilmente resueltos. Contrariamente, la subprueba motora del VMI tiene 24 ítems y permite una mayor dispersión de los puntajes. Aún con estas diferencias, ambas pruebas motoras se relacionan más allá del error de muestreo y son componentes importantes del desempeño en la prueba de pre-escritura. En segundo lugar, los ítems motores del VMI-4 tienen una heterogénea configuración y están relacionados con los ítems visomotores del subtest Copia, por lo tanto, la habilidad motora se muestra con diferentes grados de dificultad. En cambio, los ítems del subtest motor del Bender-II tienen una estructura homogénea y los ítems muestran un alto grado de semejanza en cuanto a la dificultad.

De acuerdo a nuestros resultados, se propone a) que el PE-56 muestra propiedades psicométricas básicas altamente satisfactorias, relacionándose con el desempeño motor y visomotor; b) que el PE-56 puede ser una medida destinada a evaluar la calidad del producto escrito y se complementa a las evaluaciones enfocadas en el proceso escrito según las recomendaciones de García, Rodríguez, Pacheco & Diez (2009) y Rosenblum & Livneh-Zirinski (2008); c) la evaluación de la escritura debe acompañarse con medidas independientes de habilidad visomotora, motora y visual, especialmente evaluaciones breves (por ejemplo, ver Beery, 2000; Brannigan & Brunner, 2002) que han mostrado fuertes evidencias de validez y confiabilidad en habla hispana (Merino, 2009, 2010, 2011, 2013; Merino & Benites, 2011; Merino et al., 2016); y d) la aplicación de un método holístico y global de calificación trazados puede beneficiar las evaluaciones rápidas y de despistaje de la calidad de escritura del niño. Otros aspectos de la validez y utilidad del PE-56 no han sido evaluados, y se invita a los profesionales y lectores a valorar su uso integrándolo en las evaluaciones psicopedagógicas de niños en edad preescolar. La inclusión de evaluaciones cualitativas del proceso de escritura, sin lugar a dudas, dará más información sobre la escritura del niño y debe ser implementada con el PE-56 u otra medida similar.

Finalmente, aunque el sistema de calificación propuesto no fue planeado para corresponderlo a clasificaciones explícitas del desempeño escritor, por ejemplo, los cuatro niveles de desarrollo de la calidad de la copia creados por Auzias (1977), los presentes

resultados podrían servir como un proxy de la calidad evolutiva que Auzias estableció: simulacro de escritura, copia parcial, copia legible y copia hábil. El sistema de puntuación creado tiene cuatro niveles de desempeño basados en la calidad del producto copiado y parece fácil establecer una correspondencia conceptual y práctica entre las mismas. Una limitación principal del estudio es que la consistencia entre los calificadores que usaron el método no fue calculada y ello podría poner un límite para generalizar sobre la precisión de los resultados sobre la capacidad de método para obtener puntajes reproducibles. El tamaño muestral también es una amenaza evidente para la generalización. Sin embargo, para los propósitos de este breve preliminar, los resultados son satisfactorios y permite una optimista continuidad en la investigación. Adicionalmente, dada la experiencia óptima de elevado consenso entre calificadores en medidas visomotoras (Brannigan & Bruner, 2002; Merino, 2009, 2010, 2011, 2012, 2013; Merino & Benites, 2011; Merino, et al., 2016), se puede deducir que se puede obtener un rango de elevado acuerdo con el nuevo sistema de calificación presentado para la PE-5.

REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Auzias, M. (1977). Concordances et discordances entre latéralité graphique et latéralité usuelle. *Psychologie Scolaire*, 19, 27-38.
- Beery, K. E. (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.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. doi:10.1037/0033-2909.107.2.238
- Bentler, R. M., & Wu, E. J. C. (2004). *EQS 6.1 for windows* [Statistical Program]. Encino, CA.: Multivariate Software, Inc.
- Boomsma, A., & Herzog, W. (2008). R function Swain: Correcting structural equation model fit statistics and indexes under small-sample and/or large-model conditions (Unpublished document). Recuperado de www.gmw.rug.nl/~boomsma/swain.pdf
- 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* (2nd ed.). Itasca, IL: Riverside Publishing.
- Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, 38(1), 25-56. doi:10.1207/S15327906MBR3801_2
- Cahill, S. (2009). Where does handwriting fit in? Strategies to support academic achievement. *Intervention in school and clinic*, 44(4), 223-228. doi:10.1177/1053451208328826
- Case-Smith, J., Holland, T., & Bishop, B. (2011). Effectiveness of an integrated handwriting program for first-grade student: a pilot study. *The American Journal of Occupational Therapy*, 65(6), 670-678. doi:10.5014/ajot.2011.000984
- 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(1), 209-214. doi:10.2466/PMS.90.1.209-214
- Chu, S. (1997). Occupational therapy for children with handwriting difficulties: a framework for evaluation and treatment. *British Journal of Occupational Therapy*, 60(12), 514-520. doi:10.1177/030802269706001202
- Condemarín, M., & Chadwick, M. (1991). *La escritura creativa y formal* (3ra ed.). Santiago de Chile: Andrés Bello.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cueto, F., Ramos, J. L., & Ruano, E. (2004). *PROESC: Evaluación de los procesos de escritura*. Madrid: TEA.
- Domínguez, L., & Merino, C. (2015). ¿Por qué es importante reportar los intervalos de confianza en el coeficiente Alfa de Cronbach? *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 13(2), 1326-1328.
- Feder, K. P., & Majnemer, A. (2007). Handwriting development, competency, and intervention. *Developmental Medicine & Child Neurology*, 49(4), 312-317. doi:10.1111/j.1469-8749.2007.00312.x
- Fisher, R. A. (1950). *Statistical methods for research worker*. Edinburgh, United Kingdom: Oliver and Boyd.
- Fives, H., & DiDonato-Barnes, N. (2013). Test Construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation*, 18(3), 1-7.
- García, J. N., Rodríguez, C., Pacheco, D., & Diez, C. (2009). Influencia del esfuerzo cognitivo y variables relacionadas con el TDAH en el proceso y producto de la composición escrita. Un estudio experimental. *Estudios de Psicología*, 30(1), 31-50.
- Hashempour, Z., Rostampour, M., & Behjat, F. (2015). The effect of brainstorming as a pre-writing strategy on EFL advanced learners' writing ability. *Journal of Applied Linguistics and Language Research*, 2(1), 86-99.
- Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, 16, 1-27. doi:10.1080/10705510802561279
- Jarque, C. M., & Bera A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163-172.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255-259.
- Jensen, P., Watanabe, H., & Richters, J. (1999). Who's up first? Testing for order effects in structured interview using a counterbalanced experimental design. *Journal of Abnormal Child Psychology*, 27, 439-436.
- Keogh, B., & Smith, C. (1971). Group technique and proposed scoring system for the bender-gestalt test with children. *Journal of Clinical Psychology*, 17(2), 172-175.
- Koppitz, E. M. (1984). *El Test Gestáltico de Bender: Evolución y nuevas versiones* (10a ed.). Buenos Aires: Guadalupe.
- Maldarelli, J. E., Kahrs, B. A., Hunt, S. C., & Lockman, J. J. (2015). Development, of early handwriting: visual-motor control during letter copying. *Developmental Psychology*, 51(7), 879-888. doi:10.1037/a0039424
- Merino, C. (2009). Un análisis no paramétrico de ítems de la Prueba Gestáltica del Bender Modificada para estudiantes de primaria. *Liberabit*, 15(2), 83-94.
- Merino, C. (2010). El Sistema de Calificación Cualitativa para la Prueba Gestáltica de Bender – Modificada: Estudio preliminar de sus propiedades psicométricas. *Avances en Psicología Latinoamericana*, 28(1), 63-73.
- Merino, C. (2011). Validez comparativa de tres sistemas de calificación del Test Gestáltico Visomotor de Bender. *Revista de Psicología*, 13, 90-104.
- Merino, C. (2012). Confiabilidad en el Test Gestáltico de Bender – 2da versión, en una muestra

- independiente de calificadores. *Revista de Investigación Educativa*, 30(1), 223-234.
- Merino, C. (2013). Test Gestáltico de Bender Modificado y VMI-4: Comparación de la validez incremental. *Psicoperspectivas*, 12(1), 183-204.
- Merino, C., & Allen, R. (2014). Análisis de ítems del nuevo test gestáltico visomotor de bender (2a versión). *Apuntes de Psicología*, 32, 49-56.
- Merino, C., & Benites, L. (2011). Evaluación de la confiabilidad en dos grupos de edad, usando el Sistema Cualitativo de Calificación para el Test de Bender Modificado. *Universitas Psicológica*, 10, 237-249.
- Merino, C., Allen, R., & Decker, S. (2013). Test gestáltico visomotor de bender – 2a versión. *Avaliação Psicológica*, 12(2), 275-278.
- Merino, C., Calderón-De la Cruz, G., & Manzanares, E. (2016). Estudio comparativo del acuerdo y consistencia intercalificadores en el test gestáltico visomotor de Bender 2a edición. *Revista Latinoamericana de Psicología*, 48(3), 175-182. doi:10.1016/j.rlp.2015.09.011
- Merino, C., Díaz, M., Zapata, L., & Benites, L. (2006). School psychology in Peru. In S. R. Jimerson, T. O. Oakland & P. T. Farrell (Eds.), *The Handbook International of School Psychology* (pp. 299-307). Oakland: Sage.
- Mokkink L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso J., & Patrick, D.L. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology*, 10(22). doi:10.1186/1471-2288-10-22
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Ory, D.T., & Mokhtarian, P.L. (2010). The impact of non-normality, sample size and estimation technique on goodness-of-fit measures in structural equation modeling: evidence from ten empirical models of travel behavior. *Quality & Quantity*, 44(3), 427-445. doi:10.1007/s11135-008-9215-6
- Rosenblum, S., & Livneh-Zirinski, M. (2008). Handwriting process and product characteristic of children diagnosed with developmental coordination disorder. *Human Movement Science*, 27(2), 200-214. doi:10.1016/j.humov.2008.02.011
- Rosenblum, S., Weiss, P. L., & Parush, S. (2003). Product and process evaluation of handwriting difficulties. *Educational Psychology Review*, 15(1), 41-81.
- Rosenblum, S., Weiss, P. L., & Parush, S. (2004). Handwriting evaluation for developmental dysgraphia: process versus product. *Reading and Writing*, 17(5), 433-458. doi:10.1023/B:READ.0000044596.91833.55
- Swain, A. J. (1975). *Analysis of parametric structures for variance matrices* (Unpublished doctoral dissertation). Department of Statistics, University of Adelaide, Australia.
- Toro, J., & Cervera, M. (2002). *TALE: Test de Análisis de la Lectoescritura*. Madrid: A. Machado Libros.
- Tseng, M. H., & Chow, M. K. (2000). Perceptual-Motor function of school-age children with slow handwriting speed. *The American Journal of Occupational Therapy*, 54, 83-88. doi:10.5014/ajot.54.1.83
- Urío, C., Toro, J., & Cervera, M. (2000). *Escala Magallanes de Lecto-escritura: TALE2000*. Bilbao: Grupo Albor-Cohs.
- Vinter, A., & Chartrel, E. (2010). Effects of different types of learning on handwriting movements in young children. *Learning and Instruction*, 20, 476-486. doi:10.1016/j.learninstruc.2009.07.001
- Volman, M. J., Van Schendel, B. M., & Jongmans, M. J. (2006). Handwriting difficulties in primary school children a search for underlying mechanisms. *The American Journal of Occupational Therapy*, 60(4), 451-460. doi:10.5014/ajot.60.4.451
- Weil, M. J., & Cunningham, S. J. (1994). Relationship between visuomotor and handwriting skills of children in kindergarten. *American of Occupational Therapy*, 48, 982-988.

