

Confiabilidad de la Prueba de Habilidad Intelectual DAP:IQ

Reliability of Intellectual Ability Test, DAP: IQ

Luis Honores Mendoza
Universidad Nacional Federico Villarreal
César Merino Soto
Universidad Científica del Sur

ABSTRACT

The reliability of the scores of Draw-A-Person Intellectual Ability Test for Children, Adolescents and Adults, DAP: IQ is examined with a sample of 155 children aged 6 to 11 years. Alpha coefficients were obtained from 0.63 to 0.74, the average alpha for the total sample was 0.68; It also was the difference in alpha coefficients confirmed that the low reliability of all our data with regard to the normative study. The agreement interscorer was 0.91. The results were not consistent with those reported by the authors of the test. We suggest a cautious use of this instrument for the assessment of groups with similar characteristics.

Key words: Draw-A-Person Intellectual Ability Test for Children, Adolescents and Adults; DAP: IQ; Reliability; Assessment, Intelligence.

RESUMEN

La confiabilidad de las puntuaciones de la Prueba de Habilidad Intelectual para el Dibujo de una Persona en Niños, Adolescentes y Adultos (DAP: IQ) se examina con una muestra de 155 niños y niñas de 6 a 11 años. Se obtuvieron coeficientes alfa desde 0.63 hasta 0.74, el alfa promedio para la muestra total fue de 0.68; Asimismo se realizó la diferencia de coeficientes alfa que confirmó la baja confiabilidad del total de nuestros datos con respecto al estudio normativo. El acuerdo intercalificadores fue de 0.91. Los resultados no fueron consistentes con los reportados por los autores de la prueba. Sugerimos un uso cauteloso de este instrumento para la evaluación de grupos con similares características.

Palabras clave: Prueba de Habilidad Intelectual para el Dibujo de una Persona en Niños, Adolescentes y Adultos; DAP:IQ; Confiabilidad; Evaluación; Inteligencia.

Artículo recibido/Article received: Noviembre 9 2009/November 9 2009, Artículo aceptado/Article accepted: Septiembre 18 2011/September 18 2011

Dirección correspondencia/Mail Address:

Luis Honores Mendoza, Estudiante de psicología e investigador en el área de medición psicológica y métodos cuantitativos. Email: arturhonores@aol.com

César Merino Soto, Profesor – investigador en la Facultad de Psicología de la Universidad Científica del Sur. Email: sikayax@yahoo.com.ar

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH esta incluida en PSERINFO, CENTRO DE INFORMACION PSICOLOGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET y GOOGLE SCHOLARS. Algunos de sus articulos aparecen en SOCIAL SCIENCE RESEARCH NETWORK y está en proceso de inclusion en diversas fuentes y bases de datos internacionales.

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH is included in PSERINFO, CENTRO DE INFORMACIÓN PSICOLÓGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET and GOOGLE SCHOLARS. Some of its articles are in SOCIAL SCIENCE RESEARCH NETWORK, and it is in the process of inclusion in a variety of sources and international databases.

Los Dibujos de la Figura Humana (DFH) pueden ser empleados para recopilar información relativa a las habilidades cognitivas (principalmente, la inteligencia) de un individuo. Sus interpretaciones pueden ser reportadas relativamente libres de cultura y ser útiles para las poblaciones que son generalmente difíciles de evaluar, como: niños hiperactivos o aquellos que aún no han aprendido a leer (Abell, Wood, & Liebman, 2001; Aiken, 2000). Algunas de las pruebas más conocidas que miden habilidades cognitivas a través de los DFH son las desarrolladas por Goodenough (1926), Harris (1963), Koppitz (1968), Naglieri (1988) y más recientemente Reynolds y Hickman (2004). Aunque teniendo algunas similitudes, estas pruebas se diferencian en la aplicación del DFH y sus sistemas de puntuación. Cabe señalar, que el uso y la utilidad de las pruebas del DFH han generado un debate considerable (Cox & Catte, 2000; Dumont & Smith, 1996; Hibbard, 2003; Joiner & Schmidt, 1997; Lally, 2001; Lilienfeld, Wood, & Gard, 2000; Motta, Little, & Tobin, 1993; West, 1998). Las pruebas del dibujo de la figura humana son también empleadas para fines distintos como la evaluación del desarrollo (Haeussler & Marchant, 1985; Meisels, Marsden, Wiske, & Henderson, 1997), el aprestamiento escolar (Dillard & Landsman, 1968; Ilg, Ames, Haines, y Gillespie, 1978; Weerdenburg & Janzen, 1985) y como técnica proyectiva (Bovan & Craig, 2002; Catte & Cox, 1999; Garb, Wood, & Nezworski, 2000; Lev-Wiesel & HersHKovitz, 2000; Machover, 1949).

El uso de los DFH en educación y psicología como medio para evaluar las habilidades cognitivas ha abarcado más de 100 años. Barnes (1893) y Ricci (1894) reportan algunas de las primeras observaciones vinculadas a la progresión del desarrollo de los dibujos de niños. Ellos señalaron que había distintos niveles de progresión en la complejidad y el detalle de los dibujos infantiles, a medida que los infantes avanzaban en edad. Klepsch y Logie (1982) reportaron que había un número de estudios y publicaciones sobre escolares, referidos a los dibujos infantiles que siguieron el ejemplo de los dos primeros investigadores, citados anteriormente, que también abrazaron la noción de que existía una relación entre los dibujos de los niños y sus capacidades cognitivas. Aunque estos primeros estudios y observaciones sentaron las bases para la utilización de los DFH en el estudio de la inteligencia, fue Florence Goodenough (1926) con su libro *Measurement of Intelligence by Drawings* quien presenta el primer estudio psicométrico que enlaza los dibujos y la inteligencia.

Goodenough (1926) fue la pionera en usar los DFH para evaluar la inteligencia de los niños. Ella desarrolló el Test del Dibujo de una Persona; Draw A Man Test, DAMT (Goodenough, 1926) como un medio para analizar los dibujos. El DAMT fue un test relativamente simple. Consistía en que el niño dibuje la imagen de un

hombre. El dibujo era analizado a través del examen de las relaciones entre el concepto de desarrollo e inteligencia general que se reflejaba en la calidad del dibujo. Ella asumía que la precisión y número de detalles en el dibujo indicaba el nivel de madurez intelectual del niño. El DAMT tuvo 51 ítems a calificar. Aunque el DAMT fue principalmente empleado como una medida de la habilidad cognitiva, Goodenough no descartó la posibilidad de utilizar el test como una forma de penetrar en la personalidad del niño.

Harris (1963) actualizó la investigación de Goodenough (1926) y trató de perfeccionar el DAMT. Éste, fue ampliado para incluir el dibujo de una mujer y un dibujo de sí mismo. Así, para el Test del Dibujo de Goodenough-Harris; Goodenough-Harris Drawing Test, GHDT (Harris, 1963), tres dibujos eran solicitados en lugar de uno. El GHDT fue similar al DAMT en la medida en que los detalles expresados en los dibujos de los niños reflejan sus habilidades cognitivas. Harris también se esforzó en reducir la ambigüedad del DAMT reformando el sistema de puntuación. El nuevo sistema de puntuación del GHDT contenía 73 ítems para puntuar el dibujo de un hombre, 71 ítems para puntuar el dibujo de una mujer y una escala de calidad de 12 puntos. El sistema de puntuación para el dibujo de sí mismo, no fue plenamente desarrollado. El GHDT fue el primer test basado en el dibujo de la figura humana que incluye dibujos de adolescentes en el grupo normativo.

Koppitz (1968; 1984) amplió el uso de los DFH añadiendo a la estimación de la madurez mental, la interpretación proyectiva de los dibujos de niños. A diferencia del DAMT y el GHDT, el Test del Dibujo de la Figura Humana de Koppitz, examina sólo un dibujo (al niño se le pide que dibuje a una persona) de niños de 5 hasta 12 años, luego el dibujo es evaluado por dos escalas: una, evalúa el conflicto emocional a través de 30 ítems emocionales y la otra, una escala para evaluar la maduración mental, a través también de 30 ítems evolutivos relacionados con la edad y la maduración.

Naglieri (1988) concibe un instrumento que bautiza como: Dibuja a una Persona: Un Sistema de Puntuación Cuantitativo, Draw-A-Person: A Quantitative Scoring System o simplemente, DAP. Este dispositivo fue articulado para superar las deficiencias del sistema de calificación del GHDT y de otras pruebas basadas en el DFH. A diferencia de Koppitz (1968; 1984), Naglieri considera necesaria la existencia de tres dibujos (un hombre, una mujer y el de uno mismo) en lugar de uno. El sistema de puntuación del DAP es un sistema conciso de 14 criterios que permite obtener un puntaje estandarizado de media 100 y desviación estándar 15 (CI), brindándonos comparar los resultados con otras pruebas normalizadas.

Recientemente, Reynolds y Hickman (2004) desarrollan la Prueba de Habilidad Intelectual para el Dibujo de una Persona en Niños, Adolescentes y Adultos, DAP: IQ. Un test basado en el DFH que emplea un solo dibujo y que según sus autores posee un sistema de calificación objetivo y estandarizado para estimar la habilidad intelectual. Se buscará replicar aquel supuesto, a través de este estudio, que tiene por objetivo principal brindar pruebas de la precisión de este reciente dispositivo de evaluación para su uso en niños de 6 a 11 años. Una revisión de la información científica dedicada al DAP: IQ, en habla inglesa, nos conduce al primer y único estudio empírico, realizado por Williams, Fall, Eaves y Woods-Groves (2006) que tuvo como objetivo principal explorar la consistencia interna y precisión del DAP: IQ en una muestra de 19 a 29 años; y lógicamente a la información reportada en el manual de la prueba. En español sólo se conoce de revisiones técnicas realizadas por Merino, Honores, García y Salazar (2008) y Merino, Honores y García (2009). Esto último justifica nuestro estudio a fin de aportar mayor información a la literatura concerniente al DAP: IQ.

MÉTODO

Participantes

Para el presente estudio se utilizó el tipo de muestreo no probabilístico (por conveniencia), conformando una muestra de 155 niños de ambos sexos. Siendo el de mayor número el sexo masculino con 80 (51.6 %). Los participantes provienen de tres instituciones educativas, dos de éstas son particulares y una estatal. Estas instituciones se ubican en Lima Metropolitana, dos en Lima Norte (Comas) y una en Lima Sur (Chorrillos); de esta última institución, se obtuvo mayor cantidad de participantes (42.9%). Las edades oscilan desde los 6 años (primer grado), con un total de 25 niños (16.3%), hasta los 11 años (sexto grado) con 39 participantes (25.1%), entre varones y mujeres; siendo la edad con mayor población la de 7 años (21.2 %). La información de las características demográficas de los participantes se complementa en la Tabla 1.

Sobre los datos extendidos sociodemográficos de los participantes, no se pudo obtener más datos, pero debido a evaluaciones realizadas anteriormente en dichos centros por los autores, podemos estimar algunas características relevantes se puede mencionar de forma no específica, debido que no se pudo recolectar el total de información sobre los participantes. En los colegios muestreados, cada salón es unidocente y se maneja alrededor de 30 a 40 niños por aula; los colegios particulares de nuestro estudio cuenta con más recursos materiales que los colegios públicos. Dado la zona de ubicación de los colegios, las familias de los niños típicamente alcanzan mayormente el nivel secundario, y las madres tienden a pasar más horas con el

niño que el padre ya que se ocupan del hogar y eventualmente realizan trabajos independientes; y mayoritariamente, las familias de los niños conviven con otros miembros de la familia extendida. Por lo tanto, los hogares de los niños lo integran generalmente más de tres miembros, con padres de condición civil de casados o convivientes, y pertenecientes a la clase media o menos.

Tabla 1. *Distribución de las características demográficas de los participantes*

	N	%	Media (D.E)
Sexo			
varón	80	51.6	
mujer	75	48.4	
Edad			
6 años	25	16.3	6 años 6 meses (0.29)
7 años	16	10.3	7 años 5 meses (0.30)
8 años	21	13.5	8 años 5 meses (0.30)
9 años	12	7.7	9 años 5 meses (0.29)
10 años	42	27.1	10 años 6 meses (0.30)
11 años	39	25.1	11 años 5 meses (0.27)
Grado escolar			
Primer grado	25	16.3	
Segundo grado	21	13.5	
Tercer grado	23	14.8	
Cuarto grado	27	17.4	
Quinto grado	30	19.3	
Sexto grado	29	18.7	
Total	155	100	

INSTRUMENTO

La Prueba de Habilidad Intelectual para el Dibujo de una Persona en Niños, Adolescentes y Adultos, cuyo nombre original es Draw-A-Person Intellectual Ability Test for Children, Adolescents and Adults, DAP:IQ (Reynolds & Hickman, 2004) es una prueba de rápido despistaje usada para estimar el CI por medio del dibujo de la figura humana. Fue diseñada para evaluar tanto a niños como a adultos en un rango de edad de 4 a 89 años. Para la administración, que puede ser individual o colectiva, son necesarios un lápiz, un borrador y el formulario de calificación. Aunque el material incluye una hoja titulada como hoja para la administración, sólo se requiere una hoja bond, A-4. La aplicación dura aproximadamente alrededor de 15 minutos, y el dibujo toma menos 10 minutos generalmente. Las instrucciones de aplicación requieren que el examinado dibuje una figura de sí mismo, completa y de frente, que no consista en una figura de palos o una caricatura; y se enfatiza que realice el mejor dibujo posible.

Luego, El dibujo es evaluado, teniendo en cuenta los siguientes 23 criterios: cabeza, pelo, ojos, pestañas, cejas, nariz, boca, mentón, oreja, cuello, hombros, brazos, codos, manos, torso, cintura, caderas, piernas, rodillas, tobillos, pies, ropa y accesorios. Cada uno de estos criterios es individualmente puntuado de 0 a 1 punto, 0 a 2 puntos, 0 a 3 puntos ó 0 a 4 puntos según corresponda. Cabe resaltar que la calificación se efectúa mediante la hoja de registro, que contiene los 23 criterios, y cada uno tiene al lado ejemplos gráficos que representan la creciente complejidad de cada puntuación en los ítems. Los ejemplos gráficos de calificación provienen de los dibujos incluidos en el manual y que se utilizan para el entrenamiento del sistema. La interpretación de los resultados es posible gracias a la derivación de los puntajes directos en puntajes estandarizados, especialmente se obtiene un cociente intelectual (CI), basado en una media 100 y una desviación estándar de 15. También se reportan rangos percentiles, puntajes T, puntajes z y eneatis. Adicionalmente, se ofrecen otros dos tipos de puntajes: equivalentes de edad y equivalentes de grado escolar.

Sobre la confiabilidad, Reynolds y Hickman (2004) calculan coeficientes alfa para la muestra normativa ($n = 2295$) dividiéndola según la edad, en 13 intervalos de un año para las edades de 4 a 16 años y los siguientes 9 intervalos de edad: 17 a 18, 19 a 29, 30 a 39, 40 a 54, 55 a 59, 60 a 64, 65 a 69, 70 a 74 y de 75 a 89. La media de los coeficientes alfa para todos los grupos de edad fue de 0.82 y osciló entre 0.74 a 0.87. El error estándar de medición para toda la muestra normativa fue de 4 a 5 dependiendo de cada grupo de edad. La confiabilidad test-retest fue reportada por los autores con una muestra de participantes del estado de Texas. Una muestra de 45 personas con edades comprendidas entre los 6 a 57 años que fue reevaluada en un intervalo de una semana. Sus puntajes fueron convertidos a puntuaciones estándar para superar los efectos de la edad. Los valores fueron correlacionados y se obtuvo un coeficiente de 0.86. Para hacer frente a la confiabilidad intercalificadores, Reynolds y Hickman (2004) realizaron dos estudios. En el primero, dos miembros de la editorial PRO-ED (editorial que publica la prueba) reportaron sus puntuaciones a 31 protocolos seleccionados aleatoriamente de la muestra normativa. Los miembros no tuvieron entrenamiento en el sistema de calificación del DAP: IQ, sólo utilizaron la información incluida en la hoja para la administración y puntuación del DAP: IQ. Los protocolos seleccionados representaron a los participantes con edades de 11 a 75 años. Luego, los protocolos fueron calificados y los puntajes transformados a puntuaciones estándar, resultando un coeficiente de correlación entre los grupos de puntuaciones de 0.95. En un segundo estudio, dos Ph.D. calificaron 148 protocolos cada uno. Los dibujos fueron seleccionados de una muestra de un rango de 6 a 11 años de edad, que de acuerdo con los autores, representa el rango de edad de más difícil para una

puntuación precisa. Los examinadores no recibieron alguna instrucción especial en el sistema de calificación DAP: IQ solo contaron con hoja para la administración y puntuación como guía. Los CI obtenidos se correlacionaron resultando un coeficiente de 0.91.

PROCEDIMIENTO

El análisis estadístico se dividió en tres etapas. En la primera, se estimó el acuerdo entre calificadoros a través del coeficiente de correlación producto-momento de Pearson. Cabe aclarar que los protocolos fueron corregidos por cinco calificadoros, estudiantes del curso de medición psicológica del autor secundario, los cuales calificaron 31 protocolos seleccionados aleatoriamente, y antes de estimar el acuerdo, todos en conjunto realizaron un entrenamiento en el sistema de calificación DAP: IQ. Al ser todos los examinadores estudiantes de psicología de los últimos ciclos, han realizado anteriormente entrenamientos formales en el manejo de pruebas psicológicas. En la segunda etapa, Se utilizó el coeficiente alfa de Cronbach (1951) para estimar la confiabilidad por consistencia interna en cada grupo de edad. Por último, se analizó si las diferencias entre los coeficientes alfa obtenidos en la muestra de estudio y la muestra normativa son estadísticamente significativas, para esto se utilizó la traducción al español del programa ALPHATEST (Lautenschlager, 1989) echa por Merino y Lautenschlager (2003). El programa permite realizar los cálculos que propuso Feldt (1969) para comparar dos coeficientes alfa obtenidos en muestras independientes: $w = \frac{1 - \alpha_1}{1 - \alpha_2}$. Donde α_1 y α_2 son los valores alfa en cada muestra y w se distribuye según: $(F_1)(F_2)$, con grados de libertad, respectivamente, de: $[(N_1-1), (N_1-1)(n_1-1)]$ y $[(N_2-1), (N_2-1)(n_2-1)]$, donde N: número de sujetos; n: número de ítems. Todo lo señalado anteriormente pertenece a la teoría del error muestral para el coeficiente alfa (Feldt, 1965; 1969; 1980; Hakstian & Whalen, 1976) o puede ser consultado también, como inferencias sobre α (Muñiz, 2003). Para el resto de análisis estadísticos se empleó el Statistical Package for the Social Sciences 13.0 for Windows (2005).

RESULTADOS

Acuerdo entre calificadoros.

Luego de la calificación de los 31 protocolos, los puntajes directos fueron convertidos a puntaje estandarizado (CI). La correlación entre los CI obtenidos por los cinco calificadoros fue de $r = 0.91$, el rango fue de 0.90 a 0.91.

Confiabilidad.

En la tabla 2 se observa el número de participantes y los coeficientes α obtenidos para cada grupo de edad en la

muestra de estudio y los valores correspondientes para el mismo rango de edad en la muestra normativa. En la muestra peruana, los coeficientes α oscilan desde marginalmente adecuados (0.63) hasta adecuados (0.74). Lo cual contrasta con el alto rango de coeficientes alfa obtenidos en la muestra normativa (0.77 a 0.84). Para ambas muestras, el alfa total es el valor promedio de los coeficientes hallados en cada grupo de edad.

Tabla 2. *Coefficientes alfa y número de participantes de muestras peruana y norteamericana*

Edad	Peruanos		Norteamericanos	
	α	N	α	N
6	.74	25	0.80	165
7	.63	16	0.79	189
8	.68	21	0.77	210
9	.65	12	0.82	78
10	.74	42	0.84	59
11	.64	39	0.85	88
total	.68	155	0.81	789

Comparación de coeficientes alfa

Para el estadístico de contraste que permite comparar coeficientes α obtenidos en muestras independientes (Feldt, 1969), se sostiene que: $H_0: \alpha_1 = \alpha_2$; $H_1: \alpha_1 \neq \alpha_2$. En el grupo de edad de 6 años, la diferencia de coeficientes α no resultó estadísticamente significativa ($\chi^2=0.7018$, g.l=1, $p > 0.05$). En el grupo de edad de 7 años, la diferencia de coeficientes α ; igualmente, no resultó estadísticamente significativa ($\chi^2=2.3801$, g.l=1, $p > 0.05$). La diferencia de coeficientes α no fue significativa estadísticamente en el grupo de 8 años ($\chi^2=0.994$, g.l=1, $p > 0.05$). En el grupo de 9 años, la diferencia de coeficientes α no fue significativa estadísticamente ($\chi^2=2.2694$, g.l=1, $p > 0.05$). La diferencia de coeficientes α no fue significativa estadísticamente en el grupo de 10 años ($\chi^2=2.6348$, g.l=1, $p > 0.05$). En el grupo de edad de 11 años, la diferencia de coeficientes α resultó estadísticamente significativa ($\chi^2=10.0772$, g.l=1, $p < 0.01$). Finalmente, La diferencia de coeficientes alfa totales fue estadísticamente significativa ($\chi^2=17.9122$, g.l=1, $p < 0.001$).

DISCUSIÓN

Este estudio examinó la confiabilidad de las puntuaciones de la prueba DAP: IQ. Un examen obligatorio que debiera rendir constantemente toda prueba que afirme medir un aspecto o dominio de la conducta humana. Hasta la fecha no hay reportes concernientes a las propiedades psicométricas del DAP: IQ que hayan sido publicados, excepto el estudio de Williams et al. (2006). Por tanto nuestros resultados serán discutidos con los reportados en el manual y en el estudio citado anteriormente.

La estimación de la confiabilidad de las puntuaciones para cada grupo de edad de 6 a 11 años se hizo a través del coeficiente alfa de Cronbach (1951), sin embargo la estimación de la confiabilidad del total de las puntuaciones se realizó, calculando la media aritmética, esto último por replicar el procedimiento realizado por los autores, quienes tuvieron presente al igual que nosotros, la variabilidad en cada grupo de edad, es decir, se recordó que la interpretación de la prueba esta condicionada por la edad del examinado, sobre todo en este rango de 6 a 11 años (niñez) donde el incremento del CI parece ser uno de los mayores con respecto a otras etapas del desarrollo, y que por tanto en cada edad o rango de edad se presenta un nivel de error estándar de medición. Williams et al., (2006) reportan para el total de 23 items, en una muestra de 110 participantes de 19 a 29 años, un coeficiente α de 0.82, un valor que coincide, para el mismo rango de edad, con el hallado por los autores en el estudio normativo. Es importante señalar que ambos estudios obtienen valores aceptados de medidas promedio de confiabilidad (Salvia & Ysseldyke, 2004). Sin embargo, guiados por las tradicionales clasificaciones de los niveles de confiabilidad (Nunnally y Bernstein, 1995) en nuestro estudio, lamentablemente se obtiene en su mayoría valores por debajo de los aceptables para la investigación básica, finalidad que no corresponde principalmente a este instrumento; para las edades de 6 y 10 años encontramos un coeficiente α igual a 0.74 valor que aunque aceptable, no supera el 0.80 para fines aplicados.

Vinculado a la estimación de la confiabilidad de nuestros datos, la comparación entre los coeficientes alfa de la muestra de estudio y la muestra normativa, para las edades de 6 a 10 años, encontramos una igualdad para los dos grupos examinados, esto tiene su principal explicación en la reducida cuantía de participantes en cada grupo de edad de nuestro estudio, con respecto a la muestra normativa. Este hallazgo es un fuerte motivo para replicar el estudio en muestras de mayores participantes. Para la edad de 11 años y en el total de participantes, las diferencias entre los coeficientes alfa resultan significativas lo cual confirma que de manera global nuestros resultados fueron menos confiables al compararlos con el estudio normativo.

Teniendo presente que la calificación es una de las tres principales fuentes de error claramente identificadas en el proceso de medición psicológica (Muñiz, 1998). Se realizó el acuerdo intercalificador con un entrenamiento exhaustivo, de esta forma ejercimos cierto grado de control sobre otras fuentes de error y cumplimos con el objetivo principal de evaluar la confiabilidad por consistencia interna y estimar a su vez el grado de varianza de error principalmente por parte del muestreo y heterogeneidad del contenido (Anastasi y Urbina, 1998); en otros términos, no se evaluó la confiabilidad intercalificador, sino un acuerdo

entre los calificadores que ofreció un valor de 0.91 que es igual al reportado por el segundo estudio realizado, con protocolos de niños en el mismo rango de edad por los autores, pero que a diferencia del nuestro, no se hizo con un entrenamiento previo. Es curioso que con o sin entrenamiento el acuerdo haya sido similar, esto puede explicarse quizá porque al igual que los autores coincidimos que este rango de edad (6 a 11 años) es el más difícil de puntuar con precisión, específicamente los ítems hombros, manos y tobillos fueron áreas de puntuación problemática. Estos desacuerdos parecen tener un efecto negativo sobre la confiabilidad global de las puntuaciones en este estudio ($\alpha=0.68$). En resumen, el DAP: IQ genera puntuaciones que han demostrado ser medianamente confiables sólo para las edades de 6 y 10 años.

Finalmente, es relevante señalar las limitaciones de este estudio, las cuales fueron incluir datos de una muestra restringida (sólo escolares y en un rango de 6 a 11 años), y no contar con una mayor cuantía de participantes para cada grupo etáreo (máximo 42). Los resultados no fueron consistentes por los reportados por los autores para este rango de edad, lo cual sugiere una replicación de nuestra investigación y mientras tanto, un uso cauteloso de este instrumento para la evaluación de grupos con similares características.

REFERENCIAS

- Abell, S. C., Wood, W., & Liebman, S. J., (2001). Children's human figure drawings as measures of intelligence: The comparative validity of three scoring systems. *Journal of Psychoeducational Assessment*, 19(3), 204-215.
- Aiken, R. L. (2000). *Psychological testing and assessment* (10a. ed.). Needham Heights, MA: Allen and Bacon.
- Anastasi, A. y Urbina, S. (1998). *Test psicológicos* (7a. ed.). Mexico: Prentice Hall.
- Catte, M. & Cox, M. V. (1999). Emotional indicators in children's human figure drawings. *European Child & Adolescent Psychiatry*, 8(2), 86-91.
- Cox, M. V. & Catte, M. (2000). Severely disturbed children's human figure drawings: Are they unusual or just poor drawings? *European Child & Adolescent Psychiatry*, 9(4), 301-306.
- Barnes, E. (1893). A study of children's drawings. *Pedagogical Seminary*, 2, 416-451.
- Bovan, J. & Craig, R. J. (2002). Validity of projective drawing indices of male homosexuality. *Psychological Reports*, 90(1), 175-183.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16(3), 297-334.
- Dillard, H. K. & Landsman, M. (1968). The evanston early identification scale: Prediction of school problems from the human figure drawings of kindergarten children. *Journal of Clinical Psychology*, 24(2), 227-228.
- Dumont, F. & Smith, D. (1996). Projectives and their infirm research base. *Professional Psychology: Research and Practice*, 27, 419-420.
- Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richarson reliability coefficient twenty. *Psychometrika*, 30, 357-370.
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's Alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 34, 363-373.
- Feldt, L. S. (1980). A test of the hypothesis that Cronbach Alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika*, 45, 99-105.
- Garb, H. N., Wood, J. M., & Nezworski, M. T. (2000). Projective techniques and the detection of child sexual abuse. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, 5(2), 161-168.
- Goodenough, F. (1926). *Measurement of intelligence by drawings*. Chicago: World Book.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-A-Man Test*. New York: Harcourt Brace Jovanovich.
- Hakstian, A. R. & Whalen, T. E. (1976). A k-sample significance test for independent Alpha coefficients. *Psychometrika*, 41, 219-231.
- Haeussler, I. M. & Marchant, T. (1985). *Test de Desarrollo Psicomotor: TEPSI*. Santiago: Ed. Universidad Católica.
- Hibbard, S. (2003). Acritique of Lilienfeld et al.'s (2000) the scientific status of projective techniques. *Journal of Personality Assessment*, 80(3), 260-271.
- Ilg, F. L., Ames, L. B., Haines, J., y Gillespie, C. (1978) *Tests de madurez escolar Instituto Gesell*. Buenos Aires: Paidós.
- Joiner, T. E., Jr. & Schmidt, K. L. (1997). Drawing conclusions or not from drawings. *Journal of Personality Assessment*, 69, 476-481.
- Klepsch, M. & Logie, L. (1982). *Children draw and tell: An introduction to the projective uses of children's human figure drawings*. New York: Brunner/Mazel.
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune & Stratton.
- Koppitz, E. M. (1984). *El test del dibujo de la figura humana* (10ª. ed.). Buenos Aires: Paidós.
- Lally, S. J. (2001). Should human figure drawings be admitted to court? *Journal of Personality Assessment*, 76(1), 135-139.

- Lautenschlager, G. J. (1989). Alphatest, testing for differences in values of coefficient alpha. *Applied Psychological Measurement*, 13(3) 284.
- Lev-Wiesel, R. & HersHKovitz, D. (2000). Detecting violent aggressive behavior among male prisoners through the Machover Draw-A-Person Test. *Arts in Psychotherapy*, 27(3), 171-177.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1(2), 27-66.
- Machover, K. (1949). *Personality projection in the drawing of the human figure*. Springfield, IL: Thomas.
- Meisels, S. J., Marsden, D. B., Wiske, M. S., & Henderson, L. W. (1997). *Early Screening Inventory - Revised: Examiner's manual*. Ann Arbor, Michigan: Rebus, Inc.
- Merino, C., Honores, L., García, W., y Salazar, G. (2008). Prueba de habilidad intelectual para el dibujo de una persona en niños, adolescentes y adultos, (DAP: IQ). *Repsi: Revista de Psicopedagogía*, 17, 9-13.
- Merino, C., Honores, L., y García, W. (2009) Prueba del dibujo de una persona para estimar la habilidad intelectual para en niños, adolescentes y adultos, DAP-IQ (Reynolds & Hickam, 2004). *Revista Psicopedagogía*, 26(79), 77-87.
- Merino, C. y Lautenschlager, G. J. (2003). Comparación estadística de la confiabilidad alfa de cronbach: aplicaciones en la medición educacional y psicológica. *Revista de Psicología de la Universidad de Chile*, 12(2) 129-139.
- Motta, R. W., Little, S. G., & Tobin, M. I. (1993). The use and abuse of human figure drawings. *School Psychology Quarterly*, 8, 162-169.
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 1(10), 1-21.
- Muñiz, J. (2003). *Teoría clásica de los test*. Madrid: Pirámide
- Naglieri, J. A. (1988). *Draw-a-person: A quantitative scoring system*. San Antonio, TX: The Psychological Corporation.
- Nunnally, J. C. y Bernstein, I. J. (1995). *Teoría psicométrica* (3a. ed.). México, D. F: McGraw-Hill.
- Reynolds, C. R. & Hickman, J. A. (2004). *Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults (DAP: IQ)*. Austin, TX: Pro-Ed.
- Ricci, C. (1894). L'arte de bambini. *Pedagogical Seminary*, 3, 302-307.
- Salvia, J. & Ysseldyke, J. E. (2004) *Assessment* (8a. ed.). Boston: Houghton Mifflin.
- Statistical Package for the Social Sciences 13.0 for Windows. (2005). *Software application*. Chicago: Prentice Hall.
- Weerdenburg, G. & Janzen, H. L. (1985) Predicting Grade 1 Success with a Selected Kindergarten Screening Battery. *School Psychology International*, 6(1), 13-23.
- West, M. M. (1998). Meta-analysis of studies assessing the efficacy of projective techniques in discriminating child sexual abuse. *Child Abuse & Neglect*, 22(11), 1151-1166.
- Williams, T. O., Fall, A. M., Eaves, R. E., & Woods-Groves, S. (2006). The reliability of scores for the Draw A Person Intellectual Ability Test for Children, Adolescent and Adult. *Journal of Psychoeducational Assessment*, 24(2), 137-144.