

La Validez del Dibujo de la Figura Humana como una Medida de la Madurez Intelectual en Niños Indígenas Yaquis

The Validity of the Human Figure Drawings as a Measure of Intellectual Maturity in Indigenous Yaqui Children

María Teresa Fernández Nistal¹, Frannia Aglaé Ponce Zaragoza² y Eneida Ochoa Ávila³

Resumen

Se presenta evidencia sobre la validez del Test del Dibujo de Dos Figuras Humanas – Madurativo (T2F-M) a través de la comparación de las puntuaciones en este test con las del WISC-IV, en una muestra de 151 alumnos indígenas yaquis (México). Los resultados mostraron coeficientes de correlación significativos bajos entre las puntuaciones de ambos tests, el coeficiente entre las puntuaciones del T2F-M y el índice Razonamiento Perceptivo fue más alto que los obtenidos con los otros índices, pero estas diferencias no fueron estadísticamente significativas. El análisis de correspondencia entre las clasificaciones de la capacidad intelectual de los alumnos en ambos tests mostró un porcentaje alto de discrepancias. Se concluye que el T2F-M es adecuado como una medida de *screening* de la madurez intelectual de estos alumnos en el marco de unas prácticas de evaluación psicológica que integren los resultados de este test con otros métodos, fuentes y contextos de evaluación.

Palabras clave: dibujo de la figura humana, indígenas yaquis, madurez intelectual, validez, WISC-IV

Abstract

This article was focused on the evidence related to the validity of the intellectual maturity assessment in the Two Human Figures Test (T2F-M) by comparing the scores in this test with those of the WISC-IV in a sample of 151 indigenous Yaquis students (Mexico). The results indicate a low significant correlation coefficient between the scores of both tests, the coefficient between the T2F-M scores and the Perceptual Reasoning Index was higher than those obtained with the other index scores, but these differences were not statistically significant. The correspondence analysis between the classifications of the intellectual capacity of the students in both tests displayed a high percentage of discrepancies. It can be concluded that the T2F-M is suitable as a screening measure of intellectual maturity of these students within the framework of psychological assessment practices that integrate the results of this test with other methods, sources and contexts of assessment.

Keywords: human figure drawings test, yaqui indigenous, intellectual maturity, validity, WISC-IV

Fuente de financiación: Investigación financiada por el Programa de Fomento y Apoyo a Proyectos de Investigación del Instituto Tecnológico de Sonora.

¹ Doctorado en Psicología. Profesora del Departamento de Psicología del Instituto Tecnológico de Sonora. Calle 5 de Febrero, 818 Sur. Col. Centro, 85000, Ciudad Obregón, Sonora, México. Tel.: 6444100921. Correo: teresa.fernandez@itson.edu.mx

² Maestría en Investigación Psicológica. Profesora del Departamento de Psicología del Instituto Tecnológico de Sonora, México. Correo: f.ponce@live.com

³ Doctorado en Psicología. Profesora del Departamento de Psicología del Instituto Tecnológico de Sonora, México. Correo: eochoa@itson.edu.mx

Introducción

El Test del Dibujo de Dos Figuras Humanas (T2F; Maganto & Garaigordobil, 2009a) es una prueba no verbal que evalúa el nivel madurativo-mental y la problemática emocional en niños de 5 a 12 años de edad, a partir del dibujo de dos personas; en primer lugar se solicita el dibujo de una persona completa y a continuación, el dibujo de una persona del sexo opuesto. Como una medida del funcionamiento intelectual, las autoras puntualizan que el T2F-Madurativo (T2F-M) es una prueba sensible para medir el nivel madurativo-mental, que permite una evaluación inicial (*screening*) para identificar niños con un posible retraso madurativo mental, pero no para establecer un diagnóstico de su nivel intelectual (Maganto & Garaigordobil, 2009b; Maganto & Garaigordobil, 2011).

La validez del T2F-M se determinó a partir de las correlaciones y el análisis comparativo entre las puntuaciones de la prueba con las del Test de Matrices Progresivas (TMP; Raven, 2001) del grupo de estandarización, formado por 1.122 alumnos españoles de 5 a 12 años de edad. Los resultados mostraron índices de correlación significativos en ambas figuras y en todas las edades, con valores que oscilan de bajos (.39) a moderados (.69; Maganto & Garaigordobil, 2009b). Por otro lado, el análisis comparativo entre los niveles de ejecución en el TMP y el T2F-M indicó un 60% de coincidencias en los alumnos con dificultades intelectuales, un 90% en los de capacidad intelectual media y un 55% en el nivel intelectual alto. Ningún alumno con un nivel de inteligencia alta en el TMP presentó un rendimiento bajo en el T2F-M y sólo el 3% de los que tuvieron puntuaciones altas en el T2F-M presentaron un nivel intelectual bajo en el TMP. Las autoras concluyen que una buena ejecución en el T2F-M no es probable que corresponda a un niño con limitación intelectual y que una ejecución baja en este test no es probable que corresponda a un niño con una capacidad intelectual alta (Maganto & Garaigordobil, 2009a). En un estudio posterior sobre el T2F-M, realizado con una muestra de 654 niños de diversas edades de la comunidad indígena yaqui – México – (Fernández, Tuset, & Ochoa, 2014), se encontraron correlaciones significativas entre el

T2F-M y el Test de Matrices Progresivas Coloreadas (MPC; Raven, 2001) en la mayoría de los grupos de edad, pero más bajas que las obtenidas en la muestra normativa original del test. Las correlaciones se situaron entre .21 y .38, y en el grupo de 7 años el índice de correlación fue no significativo (.18).

Existe una extensa literatura sobre la validez de los distintos sistemas del dibujo de la figura humana (DFH) que evalúan la capacidad intelectual. La mayoría de las investigaciones se basan en estudios de correlación con pruebas de inteligencia clínicas de aplicación individual, las más utilizadas han sido la Escala de Inteligencia Stanford-Binet (Thorndike, Hagen, & Sattler, 1986) y la Escala Wechsler de Inteligencia para Niños (WISC; Wechsler, 1949). Los resultados de los estudios realizados con las distintas versiones del WISC han mostrado, en general, índices de correlación significativos con valores variables, predominando los bajos y moderados. La revisión de Harris (1991) sobre el Test del Dibujo de un Hombre de Goodenough (1926) mostró correlaciones significativas con el cociente intelectual total (CIT) del WISC que se situaban entre .13 y .70. En la revisión de Scott (1981) sobre las publicaciones del sistema Goodenough-Harris (Harris, 1991) se mencionan estudios que encontraron correlaciones significativas moderadas y altas entre la escala de puntos de este sistema y los CI del WISC-R (entre .49 y .81). Investigaciones posteriores a esta revisión encontraron índices de correlación significativos bajos y moderados (rango de .29 a .63) con el CIT del WISC-R (Abell, Horkheimer, & Nguyen, 1998; Abell, Von Briesen, & Watz, 1996; Abell, Wood, & Liebman, 2001; Aikman, Belter, & Finch, 1992; Fabry & Bertinetti, 1990; Sutter & Bishop, 1986; Tramill & Edwards, 1980) y con el CIT del WISC-III (.45; Abell et al., 2001). Con respecto al sistema de Koppitz (2002), esta autora llevó a cabo estudios de correlación de su sistema con el WISC, obteniendo correlaciones significativas con el CIT que oscilaron entre moderadas (.45) y altas (.80). Estudios posteriores identificaron correlaciones significativas moderadas (.68) con el CIT del WISC (Gayton, Tavormina, Evans, & Schuh, 1974) y bajas (.28) con el CIT del WISC-R (Abell et al., 1996). La mayoría de las investigaciones realizadas con el

Dibujo de una Persona: Un Sistema de Puntuación Cuantitativo (DAP: QSS; Naglieri, 1988) también han obtenido correlaciones significativas entre bajas y moderadas con los CIT de distintas versiones de las escalas Wechsler que se sitúan entre .30 y .53 (Abell et al., 2001; Denham, 2007; Lassiter & Bardos, 1995; Willcock, Imuta, & Hayne, 2011; Wisniewski & Naglieri, 1989). La excepción a estos resultados corresponde al estudio de Nasvytiene (2007) que no identificó correlaciones significativas entre el DAP:QSS y el WISC-III. Los autores del Test de Habilidad Intelectual del Dibujo de una Persona (Reynolds & Hickman, 2004) encontraron una correlación moderada (.46) entre este sistema y el CIT del WISC-III.

La mayoría de las investigaciones que han comparado los índices de correlación obtenidos entre las puntuaciones de los sistemas del DFH y los CI de las escalas Wechsler han encontrado correlaciones más altas con el CI manipulativo que con el CI verbal (Abell et al., 1998; Abell et al., 1996; Abell et al., 2001; Fabry & Bertinetti, 1990; Gayton et al., 1974; Reynolds & Hickman, 2004; Sutter & Bishop, 1986; Tramill & Edwards, 1980), aunque esta diferencia parece no ser significativa (Wisniewski & Naglieri, 1989).

Otro aspecto que se ha abordado en los estudios de validez ha sido el análisis de la correspondencia entre las puntuaciones de los tests del DFH y los CI de la Stanford-Binet y las escalas Wechsler. Cuando este análisis se ha realizado considerando las puntuaciones promedio del grupo de participantes, los resultados indican que el nivel intelectual proporcionado por los distintos sistemas del DFH es inferior en relación a los CI obtenidos en la Stanford-Binet y las escalas Wechsler y que esta diferencia se acentúa en los niños con un nivel intelectual alto (Abell et al., 1996; Abell et al., 2001; Aikman et al., 1992; Fabry & Bertinetti, 1990; Jing, Yuan, & Liu, 1999; Lehman & Levy, 1971; Scott, 1981; Wisniewski & Naglieri, 1989). Por otro lado, las investigaciones que han realizado análisis de la correspondencia entre las clasificaciones de inteligencia individuales de los participantes proporcionadas por los sistemas del DFH y las escalas de inteligencia Wechsler han revelado un alto número de discrepancias (Aikman et al., 1992; Imuta, Scarf, Pharo, & Hayne, 2013;

Willcock et al., 2011). Aikman et al. (1992) encontraron porcentajes de desacuerdos entre las clasificaciones proporcionadas por el WISC-R y el Goodenough-Harris de 61.6% y 64.4%, y un mayor porcentaje de coincidencias en la clasificación baja de la capacidad intelectual que en las otras clasificaciones. Imuta et al. (2013) encontraron un 64% de desacuerdos con una forma abreviada de la Escala de Inteligencia de Wechsler para Preescolar y Primaria-III (WPPSI-III) y Willcock et al. (2011) un 75% con una forma abreviada del WPPSI-R.

Algunos autores han cuestionado la utilidad clínica de la técnica del DFH como una medida de la habilidad intelectual basándose en los resultados de los estudios de validez que muestran, en general, índices de correlación entre bajos y moderados con pruebas de evaluación individual de inteligencia (Aikman et al., 1992; Imuta et al., 2013; Willcock et al., 2011). Sin embargo, otros autores han interpretado estos resultados de manera diferente argumentando que el nivel relativamente bajo de las correlaciones sugiere que los DFH están evaluando otros aspectos del funcionamiento intelectual - según Knoff (2003) evalúa sólo un tipo de madurez intelectual, la cual es operacionalizada por tareas viso-motoras que requieren reconocimiento cognitivo, experiencial y ambiental de las características físicas de las personas y la habilidad para representarlas gráficamente durante una situación de test - y que esta información es útil integrada en una batería de tests amplia (Lassiter & Bardos, 1995; Sutter & Bishop, 1986). De hecho, en la práctica profesional el DFH se utiliza generalmente como un instrumento de *screening*, como una prueba más dentro de una batería de tests y como una medida de evaluación suplementaria del funcionamiento intelectual en los niños con determinadas características que impiden el uso de tests amplios como las escalas Wechsler, por ejemplo, con dificultades de atención, motivación y lenguaje (Abell et al., 2001; Flanagan & Motta, 2007; Imuta et al., 2013; Maganto & Garaigordobil, 2009b; Willcock et al., 2011).

Los estudios de validez sobre el T2F-M son escasos, únicamente se disponen de evidencias de validez de la comparación con el TMP de Raven que realizaron las autoras en la estandarización del

instrumento (Maganto & Garaigordobil, 2009a) y la de un estudio posterior realizado con una muestra de niños del grupo indígena yaqui – México – (Fernández et al., 2014) que mostró correlaciones bajas entre esta técnica del DFH y el MPC. Teniendo en cuenta que no existen investigaciones sobre la relación del T2F-M con pruebas del funcionamiento intelectual clínicas de aplicación individual, como las escalas de inteligencia Wechsler, que aportan evidencias más sólidas sobre la validez concurrente, y la contribución que representaría esta información en la adecuación de la interpretación de las puntuaciones sobre la madurez intelectual de este test en la población infantil yaqui, se plantea el objetivo de evaluar la validez del T2F-M en una muestra de alumnos de esta comunidad indígena a través de la comparación de las puntuaciones en este test con las del WISC-IV.

Basándose en los resultados de los estudios anteriores sobre este tema y las diferencias que presenta el DFH con respecto a las escalas Wechsler -concretamente que consiste en una prueba no verbal, breve y que evalúa un aspecto determinado de madurez intelectual (Knoff, 2003)-, se esperan obtener coeficientes de correlación significativos entre bajos y moderados con los CI del WISC-IV, siendo significativamente más alta la correlación con el índice de Razonamiento Perceptivo que las de los otros índices, y un porcentaje más alto de coincidencias de la clasificación del funcionamiento intelectual bajo obtenidas por los alumnos en ambos tests que de las clasificaciones del nivel medio y alto.

Método

Participantes

La muestra está formada por 151 alumnos (73 niños y 78 niñas) de cinco escuelas públicas situadas en las comunidades yaquis (Sonora, México) de 2º (n=121) y 3º (n=30) de primaria. La media de edad de los alumnos de segundo es de 7.6 ($SD=.45$) y de los de tercero de 8.6 ($SD=.27$). Todos los alumnos hablaban español, ninguno presentaba sospecha de deficiencia intelectual, estaba en tratamiento médico o psiquiátrico ni presentaba problemas visuales y/o auditivos. El tipo de muestra es no

aleatoria de participantes voluntarios puesto que el personal de la Unidad de Servicios de Apoyo a la Educación Regular (USAER) de la zona se encargó de indicar los centros escolares que aceptaron colaborar en la investigación.

Contexto. La comunidad indígena yaqui habita en una amplia zona del estado mexicano de Sonora. Antes de la conquista española vivían en pequeños agrupamientos en los márgenes del curso inferior del río Yaqui, eran agricultores seminómadas y guerreros (Spicer, 1994). Con la colonización de los misioneros jesuitas, la población yaqui se concentró en ocho pueblos que actualmente constituyen las unidades fundamentales de organización política y social de este grupo. Según el censo de población realizado en 2010 el número de habitantes fue de 31,802, de los cuales el 43.8% se comunican en lengua yaqui (Secretaría de Educación Pública [SEP], 2013).

El grado de marginación de esta comunidad es alto. Gran parte de su población se concentra en zonas rurales que no disponen de los servicios sociales básicos, con bajos ingresos económicos, vivienda inadecuada y baja escolaridad (Consejo Nacional de Población [CONAPO], 2010). En el territorio yaqui se ofrece educación escolar desde el nivel inicial hasta el superior. Los programas de estudio de las escuelas primarias se basan en los programas nacionales establecidos por la Secretaría de Educación Pública y utilizan los libros de texto en español de esta secretaría y los libros en yaqui “Jiak Noki” elaborados por el Proyecto Educativo de la Tribu Yaqui. Los indicadores educativos de la población indígena en Sonora y México, en general, están por debajo de la población no indígena, con mayores tasas de analfabetismo y menores logros educativos. En México alrededor de un tercio de niños indígenas en edad escolar (6-14 años) no asisten a la escuela y casi la mitad de la población indígena reporta como último año de estudios la primaria (Centro de Estudios Sociales y de Opinión Pública [CESOP], 2011). En la comunidad yaqui el 15% de la población de 15 años o más no tiene la educación básica, sólo el 12% cuenta con la educación media superior y el 5% superior (Instituto Nacional de Estadística y Geografía [INEGI], 2010).

Instrumentos

Se utilizaron los siguientes instrumentos: la Escala Madurativa del Test de dos Figuras Humanas (T2F-M; Maganto & Garaigordobil, 2009a) y la Escala Wechsler de Inteligencia para Niños-IV (WISC-IV; Wechsler, 2007). El T2F-M es una prueba no verbal cuyo objetivo es obtener el nivel de madurez mental del niño a partir de la elaboración del dibujo de dos figuras humanas (femenina y masculina). El procedimiento de corrección consiste en valorar si los dibujos cumplen o no los criterios de presencia de 52 ítems madurativos, lo cual permite obtener dos puntuaciones directas parciales, una en la figura masculina y otra en la femenina, y una puntuación directa total que se transforma en percentiles.

El WISC-IV es un amplio instrumento clínico de aplicación individual para evaluar la inteligencia de niños de 6 años 0 meses a 16 años 11 meses de edad, consta de 15 subpruebas que se organizan en cuatro puntuaciones que representan el funcionamiento intelectual de dominios cognoscitivos específicos (comprensión verbal, razonamiento perceptivo, memoria de trabajo y velocidad de procesamiento), así como una puntuación compuesta que representa la capacidad intelectual general del niño (CI Total). En México, la estandarización del WISC-IV se realizó con 1,100 alumnos que vivían en zonas urbanas y suburbanas de 11 estados de este país, los criterios de exclusión fueron la presencia de discapacidad física o intelectual, padecer una enfermedad y el no tener como lengua materna el español (Wechsler, 2007).

Procedimiento

Se solicitó al psicólogo de USAER de la zona escolar yaqui su apoyo para la realización de esta investigación. El psicólogo se encargó de comunicar a maestros, alumnos y padres de distintos centros escolares de la zona el objetivo de la investigación, en qué consistiría la participación de los alumnos y el uso confidencial de los resultados individuales de las pruebas. Finalmente, el psicólogo indicó los cinco centros escolares que aceptaron colaborar. El estudio cumplió con las normas éticas para la investigación con seres humanos de la APA (autorización institucional, consentimiento informado y asentimiento del alumno).

Tres psicólogas aplicaron el T2F y el WISC-IV a los alumnos de los grupos que aceptaron participar voluntariamente en la investigación en instalaciones de los centros escolares, dentro del horario escolar y durante el ciclo escolar 2015-2016. Las administraciones se realizaron según las instrucciones de los manuales de estos tests, en una o dos sesiones, dependiendo del ritmo de ejecución del alumno. El T2F se administró individualmente. Se utilizaron las normas del T2F-M del estudio de Fernández et al. (2014) y los baremos mexicanos del WISC-IV (Wechsler, 2007) para la transformación de las puntuaciones naturales a normalizadas.

Procedimiento en el análisis de los datos. El método de la investigación fue cuantitativo, con alcance correlacional y un diseño evolutivo transversal. Se aplicaron pruebas de correlación bivariada y regresión lineal para estudiar las relaciones entre las puntuaciones de los alumnos en el T2F-M y el WISC-IV. Se aplicó la prueba Z de Steiger (Steiger, 1980) de correlaciones relacionadas para comparar diferencias entre coeficientes de correlación. La comparación del funcionamiento de los participantes entre el T2F-M y el WISC-IV se realizó a través de un ANOVA de un factor con medidas repetidas y de un análisis de la correspondencia entre las clasificaciones de inteligencia individuales de los alumnos en estos tests. Esta clasificación se basó en los mismos puntos de corte (percentiles 10 y 90) de Maganto y Garaigordobil (2009a), los alumnos con un percentil 10 o inferior se clasificaron en la capacidad intelectual baja, los alumnos con un percentil entre 11 y 89 en la capacidad intelectual media y los alumnos con un percentil de 90 o superior en la capacidad intelectual superior.

El acuerdo intercalificadores se analizó a partir del coeficiente de correlación intraclase. Se seleccionaron al azar 30 protocolos del T2F-M y del WISC-IV que fueron calificados independientemente por dos psicólogos siguiendo los criterios de corrección de los manuales de estos tests. En el análisis del T2F-M se utilizó la puntuación directa total. El resultado mostró un coeficiente de correlación intraclase muy alto, $r=.944$, $p=.000$, IC 95% [.88, .97]. En el WISC-IV el acuerdo intercalificadores se calculó únicamente en los subtests de Semejanzas,

Vocabulario y Comprensión, cuya calificación requiere un juicio subjetivo, y para ello se emplearon las puntuaciones naturales totales de los subtests. Se obtuvieron coeficientes de correlación intraclase muy altos en Semejanzas, $r=.988$, $p=.000$, IC 95% [.97, .99]; Vocabulario, $r=.994$, $p=.000$, IC 95% [.98, .99] y Comprensión, $r=.976$, $p=.000$, IC 95% [.95, .98]. Teniendo en cuenta este alto nivel del acuerdo intercalificadores, el resto de los protocolos del T2F-M y del WISC-IV se dividió entre los dos evaluadores y fue calificado una sola vez. Las puntuaciones de ambos tests se ingresaron en el programa IBM SPSS Statistics 21.0 y se sometieron a un procedimiento de doble comprobación.

Resultados

Se realizó un ANOVA factorial univariante para estudiar si la puntuación directa media total en el T2F-M y la puntuación media en el CIT del WISC-IV de los alumnos difieren según las variables sexo y grado. Los resultados indicaron que no existen diferencias significativas en las puntuaciones totales medias en el T2F-M según el sexo, $F(1, 147)=1.965$, $p=.163$, $\eta_p^2=.013$; el grado, $F(1, 147) =.248$, $p=.619$, $\eta_p^2=.002$; ni efecto significativo de estas variables en su conjunto, $F(3, 147) =1.069$, $p=.364$, $\eta_p^2=.021$. En relación con el WISC-IV, tampoco se encontraron diferencias significativas según el sexo, $F(1, 147) =1.028$, $p=.312$, $\eta_p^2=.007$; grado, $F(1, 147) =.011$, $p=.915$, $\eta_p^2=.000$; ni efecto significativo de estas variables en su conjunto, $F(3, 147) =1.459$, $p=.228$, $\eta_p^2=.029$. Por consiguiente, se decidió no controlar el efecto de las variables sexo y grado en el análisis de la correlación entre las puntuaciones de los alumnos en el T2F-M y el WISC-IV.

En la Tabla 1 se presentan los coeficientes de correlación de Pearson entre el T2F-M (puntuación directa) y el WISC-IV (cocientes intelectuales). Como se puede observar, todas las correlaciones entre los índices del WISC-IV y las puntuaciones en la figura masculina, femenina y total en el T2F-M fueron positivas, estadísticamente significativas y bajas. Se obtuvieron coeficientes de correlación más altos en la figura masculina que en la femenina y la puntuación total. En la mayoría de las correlaciones se obtuvieron probabilidades inferiores al uno por ciento (.01) y presentaron intervalos de confianza que fluctúan de índices positivos muy débiles a moderados. Un análisis de regresión lineal múltiple, considerando el CIT en el WISC-IV como variable criterio y como variables predictoras las puntuaciones directas en la figura masculina, femenina y en el total del T2F-M, mostró que las distintas puntuaciones en el T2F-M son un predictor significativo del CIT en el WISC-IV, con un tamaño del efecto moderado, que explican el 15% de la varianza en este test, $R^2=.152$, $F(3,147)=8.782$, $p=.000$, $f^2 =.17$, IC 95% [.04, .32].

Los coeficientes de correlación entre las distintas puntuaciones del T2F-M y el índice RP fueron más altos que los obtenidos en CV, MT y VP (ver Tabla 1). La correlación entre la puntuación total en el T2F-M y RP fue de .36, mientras que con CV fue de .24, con MT fue de .26 y con VP de .23. La prueba Z de Steiger (Steiger, 1980) de correlaciones relacionadas mostró que no existen diferencias significativas en los índices de correlación obtenidos entre el T2F-RP vs. T2F-CV ($Z= -1.51$, $p=.130$), T2F-RP vs. T2F-MT ($Z= 1.24$, $p=.215$) ni entre el T2F-RP vs. T2F-VP ($Z= 1.39$, $p=.166$).

Tabla 1. Coeficientes de correlación de Pearson entre el T2F-M y el WISC-IV en la muestra total

| ÍNDICES WISC-IV | T2F-M | | | | | | | | |
|--------------------|----------------------|-------------|----------|---------------------|-------------|----------|-----------|-------------|----------|
| | T2F-Figura Masculina | | | T2F-Figura Femenina | | | T2F-Total | | |
| | <i>r</i> | IC 95% | <i>p</i> | <i>r</i> | IC 95% | <i>p</i> | <i>r</i> | IC 95% | <i>p</i> |
| CV | .263 | [.117,.405] | .001 | .201 | [.041,.351] | .014 | .243 | [.079,.393] | .003 |
| RP | .382 | [.239,.518] | .000 | .305 | [.178,.427] | .000 | .365 | [.218,.490] | .000 |
| MT | .278 | [.126,.416] | .001 | .240 | [.081,.384] | .003 | .269 | [.122,.413] | .001 |
| VP | .246 | [.075,.402] | .002 | .206 | [.062,.333] | .011 | .239 | [.078,.376] | .003 |
| CIT | .389 | [.235,.536] | .000 | .315 | [.157,.450] | .000 | .371 | [.192,.510] | .000 |

Nota. CV=Compresión verbal; RP=Razonamiento perceptivo; MT=Memoria de trabajo; VP=Velocidad de procesamiento; CIT=Cociente Intelectual Total.

Para comparar el funcionamiento de los alumnos en el WISC-IV y el T2F-M se calcularon los estadísticos de tendencia central y dispersión de las puntuaciones percentiles obtenidas en estos tests (ver Tabla 2). Los resultados indicaron una peor ejecución en el WISC-IV (sobre todo en CV y en el CIT) que en el T2F-M. Un ANOVA de un factor con medidas repetidas mostró que existen diferencias significativas en la puntuación media percentil obtenida en los distintos índices del WISC-IV y en el T2F-M, con un efecto grande, $F(5,146)=85.608, p=.000, \eta_p^2=.746$. El método post hoc Bonferroni mostró que la mayoría de las comparaciones entre las puntuaciones medias de los índices del WISC-IV y del T2F-M total difieren significativamente entre sí ($p=.000$), excepto las comparaciones entre CV vs. MT, CV vs. CIT y entre RP vs. VP.

En la Tabla 3 se presentan los resultados del análisis de correspondencia entre las clasificaciones del funcionamiento intelectual individual de los alumnos en el T2F-M y el WISC-IV. A nivel general, el porcentaje de coincidencias fue del 48% (22 alumnos clasificados en la categoría capacidad intelectual

baja y 50 clasificados en la categoría media en el T2F-Total y en el WISC-IV), el resto (52%) son clasificaciones discrepantes. Teniendo en cuenta las tres categorías por separado, el porcentaje mayor de coincidencias corresponde a la clasificación baja de la capacidad intelectual. La mayoría de los alumnos que se situaron en esta categoría en el T2F-Total también la presentaron en el WISC-IV (el 81%), el resto se clasificó en la categoría media (19%) y ninguno en la capacidad intelectual alta. En relación a la categoría capacidad intelectual media, existe un 45% de coincidencias, 50 alumnos se situaron en esta categoría en el T2F-Total y también en el WISC-IV, pero una frecuencia mayor de alumnos (el 55%) se situaron en la clasificación baja de capacidad intelectual en el WISC-IV. El porcentaje mayor de desacuerdos se obtuvo en la categoría de capacidad intelectual alta: del total de alumnos que se situaron en esta categoría en el T2F-Total, la mayoría (el 77%) se clasificaron en la categoría media en el WISC-IV, el 23% en la categoría baja y ninguno en la capacidad intelectual alta en este test.

Tabla 2. Estadísticos descriptivos de la puntuación percentil en las escalas del WISC-IV y en la puntuación total en el T2F-M de la muestra total

| Índices WISC-IV | <i>M</i> | <i>Mdn</i> | <i>DT</i> | IC 95% |
|-----------------|----------|------------|-----------|----------------|
| CV | 13.94 | 6 | 16.62 | [11.27, 16.61] |
| RP | 21.69 | 18 | 18.56 | [18.71, 24.68] |
| MT | 16.83 | 13 | 18.30 | [13.89, 19.78] |
| VP | 28.19 | 16 | 23.92 | [24.34, 32.04] |
| CIT | 13.04 | 8 | 15.02 | [10.62, 15.45] |
| T2F-M | 46.76 | 45 | 28.42 | [42.19, 51.33] |

Nota. CV=Compresión verbal; RP=Razonamiento perceptivo; MT=Memoria de trabajo; VP=Velocidad de procesamiento; CIT=Cociente Intelectual Total.

Tabla 3. Comparación de las categorías de capacidad intelectual obtenidas por los alumnos en el T2F-M y el WISC-IV

| WISC-IV | T2F-Total | | |
|---------------|--------------|---------------|-------------|
| | Baja (n= 27) | Media (n=111) | Alta (n=13) |
| Baja (n=86) | 22 (81%) | 61 (55%) | 3 (23%) |
| Media (n= 65) | 5 (19%) | 50 (45%) | 10 (77%) |
| Alta (n=0) | 0 | 0 | 0 |

Nota. Baja= incluye los alumnos con un percentil 10 o inferior. Media= incluye los alumnos con un percentil entre 11 y 89. Alta= incluye los alumnos con un percentil de 90 o superior.

Discusión

El objetivo de este estudio consistió en evaluar la validez del T2F-M en una muestra de alumnos de 7 y 8 años de la comunidad indígena yaqui a través de la comparación de las puntuaciones en este test con el WISC-IV. No existen estudios que hayan comparado este sistema del DFH con una prueba clínica de aplicación individual de la capacidad intelectual, como las escalas Wechsler, por lo que los resultados que se han obtenido aportan evidencias de la validez de este sistema de la figura humana. Los principales resultados mostraron coeficientes de correlación significativos bajos entre las puntuaciones en el T2F-M y el WISC-IV, que el coeficiente de correlación T2F-M vs. RP fue más alto que los obtenidos con los otros índices, aunque estas diferencias no fueron significativas, y un porcentaje alto de clasificaciones discrepantes de la capacidad intelectual de los alumnos entre ambos tests.

Los resultados de la relación entre las distintas puntuaciones de los alumnos en el T2F-M y los CI del WISC-IV mostraron índices de correlación significativos bajos, que oscilan de .389 a .201, con un riesgo de error en la mayoría de las correlaciones inferior al uno por ciento. Estos resultados son consistentes con los de la mayoría de los estudios anteriores realizados con otros sistemas del DFH y versiones del WISC que encontraron correlaciones significativas con valores variables, predominando los bajos y moderados (Abell et al., 1998; Abell et al., 1996; Abell et al., 2001; Aikman et al., 1992; Denham, 2007; Fabry & Bertinetti, 1990; Gayton et al., 1974; Koppitz, 2002; Lassiter & Bardos, 1995; Reynolds & Hickman, 2004; Sutter & Bishop, 1986; Tramill & Edwards, 1980; Willcock et al., 2011; Wisniewski & Naglieri, 1989). Algunos autores han interpretado los coeficientes de correlación bajos y moderados entre los DFH y las escalas Wechsler como una muestra de debilidad del DFH para hacer inferencias sobre la habilidad intelectual (Aikman et al., 1992; Imuta et al., 2013; Willcock et al., 2011). Otros autores, por el contrario, han considerado que estos resultados pueden indicar que los DFH están evaluando otros aspectos del funcionamiento intelectual (Lassiter & Bardos, 1995; Sutter & Bishop, 1986;

Wisniewski & Naglieri, 1989). En el presente estudio se considera que los resultados obtenidos aportan evidencia de validez concurrente de las puntuaciones del T2F-M (correlaciones significativas con el WISC-R con riesgo de error pequeño, con intervalos de confianza que fluctúan de correlaciones positivas muy débiles a moderadas y los resultados del análisis de regresión lineal múltiple que indicaron que las distintas puntuaciones del T2F-M son un predictor significativo, con un tamaño del efecto moderado, del CIT en el WISC-IV). Los coeficientes de correlación bajos y el porcentaje relativamente pequeño de varianza (15%) que explica la puntuación del T2F-M en el CIT del WISC-IV indican que este sistema del DFH está evaluando determinadas funciones que son comunes con el WISC-IV, además de otros aspectos específicos de la capacidad intelectual, entendida como un constructo complejo y multidimensional.

El coeficiente de correlación entre la puntuación total del T2F-M y el índice RP fue más alto (.365) que los obtenidos con los índices de CV (.243), MT (.269) y VP (.239), sin embargo estas diferencias no fueron estadísticamente significativas. La mayoría de los estudios anteriores encontraron correlaciones más altas entre las puntuaciones del DFH y el CI manipulativo de las escalas Wechsler que con el CI verbal, pero no analizaron si esta diferencia entre las correlaciones era estadísticamente significativa (Abell et al., 1998; Abell et al., 1996; Abell et al., 2001; Fabry & Bertinetti, 1990; Gayton et al., 1974; Reynolds & Hickman, 2004; Sutter & Bishop, 1986; Tramill & Edwards, 1980). Los autores que analizaron este aspecto, Wisniewski y Naglieri (1989), obtuvieron una correlación más alta entre el DAP: QSS y el CI manipulativo del WISC-R que con el CI verbal y no encontraron diferencias significativas entre ambos índices de correlación. Estos autores concluyen que posiblemente este sistema del DFH evalúe aspectos más generales de la habilidad intelectual y no específicos, basados en un modelo de inteligencia verbal o no verbal como el WISC-R. Los resultados obtenidos en el presente estudio no permiten comprobar el supuesto planteado sobre la mayor relación entre el T2F-M y el índice de RP debido a su naturaleza no verbal. El hecho de que se hayan obtenido correlaciones

significativas entre las puntuaciones del T2F-M y el resto de los índices del WISC-IV y el CIT indica que el funcionamiento en el T2F-M está relacionado no exclusivamente con el razonamiento perceptivo, sino también con otros aspectos generales de la capacidad intelectual que evalúa el WISC-IV.

La comparación entre el funcionamiento de los alumnos en el T2F-M y el WISC-IV mostró una peor ejecución en este último test. La media en puntuación percentil obtenida por la muestra total de alumnos en el T2F-M se situó en un valor promedio ($M=46.76$, $DT=28.42$), en cambio las puntuaciones medias percentiles obtenidas en los índices del WISC-IV fueron significativamente más bajas, sobre todo el índice de CV ($M=13.94$ y $DT=16.62$) y el CIT ($M=13.04$ y $DT=15.02$). Estos resultados no coinciden con los de estudios anteriores que, en general, han encontrado puntuaciones más altas de los niños en las distintas versiones de las escalas Wechsler y la Stanford-Binet, que en los sistemas del DFH (Abell et al., 1996; Abell et al., 2001; Aikman et al., 1992; Fabry & Bertinetti, 1990; Jing et al., 1999; Lehman & Levy, 1971; Scott, 1981; Wisniewski & Naglieri, 1989). Esta discrepancia con los estudios anteriores puede explicarse por la falta de representación del grupo indígena ya que en las normas del WISC-IV que se utilizaron y la inadecuación cultural de las tareas de esta prueba para evaluar su capacidad intelectual. Con respecto a las normas, los estudios anteriores se realizaron con muestras de alumnos estadounidenses – para los que fueron creados estas pruebas – y se utilizaron baremos representativos de las características socioculturales de esos niños. En la presente investigación los alumnos pertenecen a un grupo indígena para los que, en muchas ocasiones, no hay disponibles normas representativas de las pruebas psicológicas, como es el caso del WISC-IV, cuyos únicos baremos en México (Wechsler, 2007) se obtuvieron de una muestra de estandarización que no incluye este grupo indígena ni otros del país. Por otro lado, el sesgo de los tests tradicionales de inteligencia (como las escalas Wechsler) hacia tareas que requieren habilidades que predominan en el contexto escolar de la cultura occidental de clase media cuestiona su adecuación para evaluar la capacidad

intelectual de los niños indígenas o de otros grupos minoritarios (Anastasi & Urbina, 1998; Dana, 1984). Estos niños, en general, tienen experiencias de aprendizaje y valores diferentes a los de esa cultura, por lo que algunas de las tareas que propone el WISC-IV no tienen mucha relevancia para ellos y su experiencia escolar con las habilidades que se requieren para su solución es escasa.

Los resultados del análisis de correspondencia entre las clasificaciones de la inteligencia individual de los alumnos en el T2F-M y el WISC-IV se vieron afectados por la ausencia de representación de los yaquis en la baremación utilizada del WISC-IV. Se encontró un alto porcentaje de clasificaciones discrepantes entre ambos tests (el 52%) que corresponden fundamentalmente a la clasificación alta y media de la capacidad intelectual y un porcentaje mayor de coincidencias en la clasificación baja de la capacidad intelectual (22 de los 27 alumnos clasificados en el T2F-M en este grupo también fueron clasificados en este nivel en el WISC-IV). Probablemente esta tendencia de los resultados se deba a la subestimación de la capacidad intelectual de los alumnos en el WISC-IV (la mayoría de ellos se situaron en la capacidad intelectual baja y ninguno en la alta). Teniendo en cuenta esta situación, la comparación de los resultados obtenidos con los de estudios anteriores que utilizaron como prueba criterio las escalas Wechsler indica coincidencias (Aikman et al. 1992; Imuta et al., 2013; Willcock et al., 2011). Por ejemplo, Aikman et al. (1992) encontraron un porcentaje alto de desacuerdos (61.6% y 64.4%) entre las clasificaciones proporcionadas por el WISC-R y el Goodenough-Harris y un mayor porcentaje de coincidencias en la clasificación baja de la capacidad intelectual que en las otras clasificaciones. Sin embargo, con el TMP Maganto y Garaigordobil (2009a) encontraron menos desacuerdos en la comparación de los niveles de ejecución obtenidos por los alumnos en este test y el T2F-M (entre el 10% y el 45%). Posiblemente, el hecho de que el TMP sea una prueba no verbal y breve, como las pruebas del dibujo de la figura humana, haya influido en este incremento de las coincidencias en las clasificaciones obtenidas por ambos tests.

La principal limitación de este estudio consistió en la falta de representación de la población infantil yaqui en las normas disponibles en México del WISC-IV, lo cual no permitió establecer conclusiones sobre el análisis de la correspondencia entre las clasificaciones del funcionamiento intelectual de estos alumnos en el T2F-M y el WISC-IV, que hubieran aportado información sobre la utilidad predictiva de este sistema del DFH en las distintas clasificaciones de la capacidad intelectual. Al margen de esto, la comparación de las puntuaciones de los alumnos de la muestra en ambos tests pone de manifiesto las consecuencias negativas del uso de estas normas, que subvaloran la capacidad intelectual de los niños yaquis, y la necesidad de investigaciones dirigidas a la adaptación y obtención de normas específicas de pruebas de inteligencia clínicas de aplicación individual en esta población. Otra debilidad de la investigación hace referencia a las características de la muestra que limitan la generalización de los resultados obtenidos.

En conclusión, los resultados de esta investigación confirman que existe una correlación significativa, aunque baja, entre las puntuaciones de los alumnos yaquis en el T2F-M y el WISC-IV, lo cual significa que este sistema del DFH está midiendo ciertas funciones de la inteligencia que también evalúa el WISC-IV, además de otros aspectos. La escasez de adaptaciones y normas específicas de las pruebas clínicas de aplicación individual más usadas, como el WISC-IV, en las poblaciones indígenas de México hace particularmente eficaz el uso de pruebas no verbales y breves, como el DFH, en estos contextos puesto que la tarea que proponen puede ser más pertinente y familiar para estos niños y, por otro lado, porque la adaptación y obtención de normas resulta menos costosa que las de las pruebas clínicas de aplicación individual. Teniendo en cuenta que existen baremos específicos de la técnica del DFH en esta población indígena (Fernández et al., 2014; Fernández, Tuset, & Ross, 2015) y los resultados obtenidos de validez concurrente del T2F-M, se considera que este sistema del DFH es adecuado como una medida de *screening* de la madurez intelectual de los alumnos de la comunidad indígena yaqui de 7 y 8 años en el marco de unas

prácticas de evaluación psicológica que integren los resultados de este test con otros métodos, fuentes y contextos de evaluación.

Referencias

- Abell, S. C., Horkheimer, R., & Nguyen, S. E. (1998). Intellectual evaluations of adolescents via Human Figure Drawings: An empirical comparison of two methods. *Journal of Clinical Psychology, 54*(6), 811-815. doi:10.1002/(SICI)1097-4679(199810)54:6<811::AID-JCLP8<3.0.CO;2-J
- Abell, S. C., Von Briesen, P. D., & Watz, L. S. (1996). Intellectual evaluations of children using Human Figure Drawings: An empirical investigation of two methods. *Journal of Clinical Psychology, 52*(1), 67-74. doi: 10.1002/(SICI)1097-4679(199601)52:1<67::AID-JCLP9>3.0.CO;2-T
- 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*, 204-215. doi: 10.1177/073428290101900301
- Aikman, K. G., Belter, R. W., & Finch, A. J. (1992). Human figure drawings: Validity in assessing intellectual level and academic achievement. *Journal of Clinical Psychology, 48*, 114-120. doi:10.1002/1097-4679(199201)48:1<114::AID-JCLP2270480116>3.0.CO;2-Y
- Anastasi, A., & Urbina, S. (1998). *Tests psicológicos* (7ª ed.). Naucalpan de Juárez, Edo. de México: Pearson.
- Centro de Estudios Sociales y de Opinión Pública (CESOP) (2011). *Situación de los indígenas, n° 16. Descripción sociodemográfica de la población hablante de lengua, autoadscrita como indígena y el resto de la población, a partir de los datos del Censo de Población y Vivienda 2010*. México D.F.: CESOP. Recuperado de <http://www3.diputados.gob.mx/camara/content/view/full/75718>
- Consejo Nacional de Población (CONAPO). (2010). *Índice de marginación por localidad 2010*. México: CONAPO. Recuperado de <http://www.conapo.gob.mx>.

- Dana, R. H. (1984). Intelligence Testing of American Indian Children: Sidesteps in quest of ethical practice. *White Cloud Journal*, 3(3), 35-43.
- Denham, S-K. M. (2007). Quantitative analysis of drawings in a referred sample of African American and Latino children. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 67(9-B), 5396.
- Fabry, J. J., & Bertinetti, J. F. (1990). A construct validation study of the human figure drawing test. *Perceptual and Motor Skills*, 70, 465-466. doi: <https://doi.org/10.2466/pms.1990.70.2.465>
- Fernández, M. T., Tuset, A. M., & Ochoa, E. (2014). Un estudio normativo de los ítems madurativos del Test del Dibujo de Dos Figuras Humanas en niños indígenas Yaquis. *Interdisciplinaria. Revista de Psicología y Ciencias Afines*, 31(2), 297-322. doi:10.16888/interd.2014.31.2.7
- Fernández, M. T., Tuset, A. M., & Ross, G. P. (2015). Un estudio normativo de los ítems evolutivos del Test del Dibujo de la Figura Humana en niños indígenas Yaquis. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 39, 77-90.
- Flanagan, R., & Motta, R. W. (2007). Figure drawings: a popular method. *Psychology in the School*, 44, 257-270. doi: 10.1002/pits.20221
- Gayton, W. F., Tavormina, J., Evans, H. E., & Schuh, J. (1974). Comparative validity of Harris' and Koppitz' scoring systems for human-figure drawings. *Perceptual and Motor Skills*, 39, 369-370. doi: 10.2466/pms.1974.39.1.369
- Goodenough, F. L. (1926). *Measurement of intelligence by drawings*. New York, United States: Harcourt, Brace and World.
- Harris, D. B. (1991). *El Test de Goodenough, revision, ampliación y actualización*. Barcelona, España: Paidós.
- Imuta, K., Scarf, D., Pharo, H., & Hayne, H. (2013). Drawing a close to the use of human figure drawings as a projective measure of intelligence. *PLoS ONE*, 8(3), e58991. doi:10.1371/journal.pone.0058991
- Instituto Nacional de Estadística y Geografía (INEGI) (2010). Censo de población y vivienda 2010. Recuperado de <http://www3.inegi.org.mx/rnm/index.php/catalog/71>
- Jing, J. Yuan, C., & Liu, J. (1999). Study of human figure drawings in learning disabilities. *Chinese Mental Health Journal*, 13(3), 133-134.
- Knoff, H. M. (2003). Evaluation of projective drawings. En C. R. Reynolds & R. W. Kamphaus (Eds.). *Psychological and Educational assessment of children* (2a ed.) (pág. 91-158). New York, United States: Guilford Press.
- Koppitz, E. M. (2002). *El dibujo de la figura humana en los niños* (12a. ed.). Buenos Aires, Argentina: Guadalupe.
- Lassiter, K. S., & Bardos, A. N. (1995). The relationship between young children's academic achievement and measures of intelligence. *Psychology in the Schools*, 32, 170-177. doi:10.1002/1520-6807(199507)32:3<170::AID-D-PITS2310320303>3.0.CO;2-K
- Lehman, E. B., & Levy, B. I. (1971). Discrepancies in estimates of children's intelligence: WISC and human figure drawings. *Journal of Clinical Psychology*, 27(1), 74-76. doi:10.1002/1097-4679(197101)27:1<74::AID-JCLP2270270114>3.0.CO;2-2
- Maganto, C., & Garaigordobil, M. (2009a). *Test del dibujo de dos figuras humanas (T2F)*. Madrid, España: TEA Ediciones.
- Maganto, C., & Garaigordobil, M. (2009b). El diagnóstico infantil desde la expresión gráfica: el Test de Dos Figuras Humanas (T2F). *Clínica y Salud*, 20, 237-248.
- Maganto, C., & Garaigordobil, M. (2011). Indicadores emocionales complementarios para la evaluación emocional del test del dibujo de dos figuras humanas (T2F). *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 31, 73-95.
- Naglieri, J. A. (1988). *DAP: Draw a person: A quantitative Scoring System Manual*. San Antonio, United States: The Psychological Corporation.
- Nasvytiene, D. (2007). The analysis of psychometric properties of the Human Figure Drawings Test. *Psichologija*, 36, 61-73.

- Raven, J. C. (2001). *Matrices Progresivas (SPM, APM y CPM)* (3ª edición). Madrid, España: TEA Ediciones.
- Reynolds, C. R., & Hickman, J. A. (2004). *Draw-A-Person Intellectual Ability Test for children, adolescents, and adults (DAP: IQ)*. Austin, United States: Pro-Ed.
- Secretaría de Educación Pública (SEP) (2013). *Guía para el maestro de la asignatura estatal, Campo 4: Jiak noki into jiak yo'olut'uria Lengua y Cultura indígena del pueblo Yaqui, del Programa de Estudio 2011 de Educación Básica. Educación Básica, Secundaria*. Hermosillo, México: SEP.
- Scott, L. H. (1981). Measuring intelligence with the Goodenough-Harris Drawing Test. *Psychological Bulletin*, 89(3), 483-505. doi:10.1037/0033-2909.89.3.483
- Spicer, E. (1994). *Los Yaquis: Historia de una cultura*. México: Universidad Autónoma de México.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251. doi:10.1037/0033-2909.87.2.245
- Sutter, E. G., & Bishop, P. C. (1986). Further investigation of the correlations among the WISC-R, PIAT, and DAM. *Psychology in the Schools*, 23, 365- 367. doi:10.1002/1520-6807(198610)23:4<365::AID-PITS2310230409>3.0.CO;2-F
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Guide for administering and scoring, the Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago, United States: Riverside Publishing.
- Tramill, J. L., & Edwards, R. P. (1980). Comparison of the Goodenough-Harris Drawing Test and the WISC-R for children experiencing academic difficulties. *Perceptual and Motor Skills*, 50, 543-546. doi:10.2466/pms.1980.50.2.543
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. San Antonio, United States: The Psychological Corporation.
- Wechsler, D. (2007). *WISC-IV: Escala Wechsler de Inteligencia para Niños-IV: Manual técnico*. México D.F.: El Manual Moderno.
- Willcock, E., Imuta, K., & Hayne, H. (2011). Children's human figure drawings do not measure intellectual ability. *Journal of Experimental Child Psychology*, 110, 444-452. doi:10.1016/j.jecp.2011.04.013
- Wisniewski, J. J., & Naglieri, J. A. (1989). Validity of the Draw-a-Person: A Quantitative Scoring System with the WISC-R. *Journal of Psychoeducational Assessment*, 7, 346 -351. doi: 10.1177/073428298900700408