

Validez Estructural de la Escala Wechsler de Inteligencia para Niños-IV en Estudiantes Indígenas de México

Structural Validity of the Wechsler Intelligence Scale for Children–Fourth Edition among Mexico’s Indigenous Students

María Teresa Fernández Nistal¹, Ana María Tuset Bertran² y Guadalupe de la Paz Ross Arguelles³

Resumen

En este artículo se presenta evidencia sobre la validez estructural de la WISC-IV en una muestra de 151 estudiantes indígenas yaquis (México) de 7 y 8 años de edad, a través de un estudio de la matriz de intercorrelaciones de las pruebas y de un análisis factorial exploratorio y otro confirmatorio. Los resultados mostraron coeficientes de correlación entre las pruebas bajos y moderados, el análisis factorial exploratorio indicó algunas diferencias con respecto al modelo de los cuatro factores y el análisis factorial confirmatorio respaldó el modelo jerárquico de los cuatro factores de primer orden y un factor de inteligencia general de segundo orden, el cual explica la mayor cantidad de la varianza común (76.8%) y total (35.8%). Se concluye que esta escala es útil para evaluar la inteligencia de este alumnado en el marco de unas prácticas de evaluación que consideren otros métodos, fuentes y las características específicas del contexto de evaluación.

Palabras clave: WISC-IV, análisis factorial, validez, evaluación psicológica, indígenas yaquis

Abstract

In this paper we present evidence about the structural validity of the WISC-IV in a sample of 151 indigenous Yaqui students (Mexico) between the ages of 7 and 8 years old through an analysis of the test’s intercorrelations matrix and of an exploratory and confirmatory factor analysis. The analysis of the matrix showed low to moderate correlation coefficients between the tests. The exploratory factor analysis indicated some differences on the model of the four WISC-IV factors. The confirmatory factor analysis supported the hierarchical model of the four first-order factors and a second-order general intelligence factor. This accounted for the greater amount of the common (76.8%) and total variance (35.8%). In conclusion this scale is useful to assess the intelligence of these students in the framework of assessment practices that consider other methods, sources, and the specific characteristics of the context.

Keywords: WISC-IV, factor analysis, validity, psychological assessment, yaqui indigenous

Agradecimientos: Esta investigación ha sido 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 Doctora en Psicología. Profesora emérita del Departamento de Psicología Clínica y Psicobiología; Facultad de Psicología; Universidad de Barcelona. Pg. de la Vall d'Hebron, 171, Edificio de Ponent 08035, Barcelona, España. Correo: atuset@ub.edu

³ Doctorado en planeación estratégica para la mejora del desempeño. Profesora del Departamento de Psicología del Instituto Tecnológico de Sonora, México. Correo: guadalupe.ross@itson.edu.mx

Introducción

Las escalas Wechsler son los tests estandarizados de administración individual más utilizados en la evaluación de la inteligencia de los niños y las niñas en Estados Unidos y han sido constantemente traducidos a otras lenguas y utilizados en diversos países (Prifitera, Saklofske, Weiss, & Rolfhus, 2005). En México, la última versión de las escalas Wechsler para Niños adaptada y estandarizada es la Escala Wechsler de Inteligencia para Niños-IV (WISC-IV; Wechsler, 2007). La estandarización se realizó con una muestra de 1 100 estudiantes de educación básica que vivían en zonas urbanas y suburbanas de diferentes estados de ese país. En el manual técnico de esta versión no se especifica sobre la inclusión de grupos étnicos del país, no obstante entre los criterios de exclusión constan el alumnado que no tenía como lengua materna el español y el que provenía del medio rural, lo cual indica que probablemente no se consideraron los grupos indígenas. En la actualidad, estos grupos constituyen una población de aproximadamente 12 millones de personas, el 10% de la población total de México, distribuidas en más de 67 grupos con diversas características étnicas, lingüísticas, culturales y regionales. La mayoría de esta población vive en municipios indígenas (Comisión Nacional para el Desarrollo de los Pueblos Indígenas, 2017).

La Dirección General de Educación Indígena de México recomienda la aplicación de la WISC-IV en la evaluación del funcionamiento intelectual de los niños y las niñas indígenas y señala la importancia de la aplicación de pruebas sin sesgo cultural (Secretaría de Educación Pública, 2012). Para la evaluación del sesgo cultural de las pruebas psicológicas se requiere evidencia empírica sobre su validez de constructo, es decir, la determinación del grado en el que puede afirmarse que mide el mismo constructo en diversos grupos étnicos (Sattler, 2010). Existen diferentes aspectos a considerar en la validez de constructo (contenido, sustantivo, estructural, generalización, etc.; Messick, 1995). El aspecto más analizado en los estudios sobre el sesgo cultural de los tests de inteligencia ha sido el estructural (Valencia & Suzuki, 2001). Este aspecto se establece cuando la estructura interna

de un instrumento es consistente con lo que se conoce sobre la estructura del constructo que se mide (Messick, 1995). El método que se ha utilizado con más frecuencia para analizar la consistencia de la estructura interna de las escalas Wechsler en diferentes poblaciones ha sido el análisis factorial (Valencia & Suzuki, 2001).

La WISC-IV (Wechsler, 2003) está organizada en cuatro índices teóricos y de puntuación: el índice de Comprensión verbal (CV), el índice de Razonamiento perceptual (RP), el índice de Memoria de trabajo (MT) y el índice de Velocidad de procesamiento (VP); y en una medida de funcionamiento intelectual general (CIT). Cada uno de los cuatro índices está compuesto por pruebas que miden determinadas capacidades. Esta estructura interna está fundamentada en la teoría de la inteligencia de Cattell-Horn-Carroll (Carroll, 1997), en los resultados de las investigaciones sobre evaluación intelectual, desarrollo cognoscitivo, neurociencia cognitiva y en los estudios analítico factoriales sobre las capacidades cognitivas (Flanagan & Kaufman, 2004; Wechsler, 2003).

La evidencia de validez basada en la estructura interna de la WISC-IV (Wechsler, 2003) se obtuvo a través del estudio de las intercorrelaciones de las puntuaciones en las pruebas y de análisis factoriales exploratorios y confirmatorios, que se realizaron con la muestra de estandarización formada por 2200 personas estadounidenses de 6 a 16 años y 11 meses de edad.

El estudio de intercorrelación de la WISC-IV se basó en la metodología de interpretación de los patrones de correlaciones en una matriz multirrasgo-multimétodo de Campbell y Fiske (1959). Los resultados de la matriz de la muestra total de estandarización indicaron que todas las correlaciones interpruebas fueron estadísticamente significativas; que las pruebas de cada índice correlacionaron en mayor grado unas con otras que con respecto al resto de las pruebas, aunque también presentaron correlaciones moderadas con algunas pruebas de otros índices con las que comparten determinadas capacidades. Por otro lado, los resultados de las matrices de intercorrelación de los grupos de 7 y 8 años de edad mostraron un patrón semejante al de la muestra total, pero con algunas excepciones. En

general, las correlaciones fueron más bajas y algunas no significativas, la correlación más alta fue entre Vocabulario y Semejanzas. Los coeficientes de correlación entre las pruebas de CV fueron los más altos, aunque también estas pruebas presentaron coeficientes moderados con las de RP y MT. En el resto de los índices se obtuvieron correlaciones bajas y moderadas entre las pruebas que los componen y presentaron correlaciones semejantes con pruebas de otros índices con las que comparten capacidades. Entre las pruebas principales, Claves fue la que presentó la correlación más baja con el resto de las pruebas.

El objetivo de los estudios analítico-factoriales fue comprobar la estructura de los cuatro factores (CV, RP, MT y VP). Los resultados del análisis factorial exploratorio (AFE) realizado con el total de la muestra fueron consistentes con el modelo factorial pronosticado; la carga principal de cada prueba residió claramente en su factor correspondiente. No obstante, en el grupo de 6 y 7 años de edad la prueba Conceptos con dibujos obtuvo cargas semejantes en los factores de CV y RP, y Semejanzas, a pesar de que su carga más alta se situó en el factor CV, presentó una carga secundaria superior a .20 en RP (Wechsler, 2003). Sattler (2010) presentó los resultados de los AFE por cada grupo de edad de la muestra de estandarización, a diferencia del manual técnico de la WISC-IV que lo hace agrupando distintas edades. En general, los resultados que presenta este autor indicaron que el modelo de los cuatro factores se reproduce, pero que existe un patrón diverso según la edad. En los grupos de edad de 7 y 8 años las pruebas que mostraron un patrón inconsistente fueron Semejanzas, Conceptos con dibujos, Vocabulario, Matrices, Búsqueda de símbolos, Figuras incompletas, Información, Aritmética y Palabras en contexto (Pistas).

En cuanto a los análisis factoriales confirmatorios (AFC) realizados con las pruebas principales de la WISC-IV, se sometieron a contrastación diferentes modelos de uno a cuatro factores de primer orden. Los resultados mostraron que el modelo de los cuatro factores oblicuos fue el que presentó el mejor ajuste a los datos observados (Wechsler, 2003). Sin embargo, en estos análisis no se consideró el factor general de inteligencia (g) de segundo orden para

comprobar y describir la estructura multinivel de la escala (Watkins, 2006). Posteriormente, Watkins (2006) analizó la estructura factorial de segundo orden en la muestra de estandarización de la WISC-IV y encontró que el factor de inteligencia general explicó la mayor cantidad de varianza común (71.3%) y total (38.3%). El AFC realizado por Keith, Goldenring, Taub, Reynolds y Kranzler (2006) mostró que el modelo jerárquico de segundo orden se ajustaba bien a los datos de la muestra normativa de la WISC-IV.

El análisis de la estructura interna de la adaptación de la WISC-IV en México (Fina, Sánchez-Escobedo, & Hollingworth, 2012) se realizó a través de un AFC y de un estudio de la matriz de intercorrelaciones de las pruebas. Los resultados del AFC mostraron un buen ajuste del modelo jerárquico de segundo orden a los datos observados. Las cargas estandarizadas de los factores de primer orden hacia las pruebas fueron altas, excepto la carga de VP hacia Claves que fue moderada. Las cargas del segundo factor (g) hacia los factores de primer orden fueron también altas, la más baja correspondió a la de VP. Por otro lado, el estudio de la matriz de intercorrelaciones mostró patrones de correlación semejantes a los observados en la estandarización estadounidense. Todos los coeficientes de correlación fueron positivos y superiores a cero, la correlación más alta fue entre Vocabulario y Comprensión, la más baja entre Claves y Conceptos con dibujos, y las correlaciones entre las pruebas que componen cada índice fueron, en su mayoría, más altas entre sí (sobre todo CV) que entre aquellas que forman otros índices, sin embargo también encontraron correlaciones moderadas entre pruebas de diferentes índices.

Las investigaciones sobre la validez estructural de las escalas Wechsler para Niños en grupos indígenas americanos presentan resultados discrepantes. Las primeras investigaciones se llevaron a cabo con la escala Wechsler de Inteligencia para Niños-Revisada (WISC-R; Wechsler, 1974) y se utilizaron técnicas de AFE. Los resultados no mostraron una reproducción totalmente semejante a la estructura interna de esta escala (McShane & Plas, 1982; Reschly, 1978; Zarske, Moore, & Peterson, 1981). Reschly (1978) comparó la estructura factorial de la WISC-R de una muestra de alumnado de primaria

y secundaria de distintos grupos étnicos estadounidenses, uno de los cuales eran 240 indígenas pápagos. Los resultados mostraron que el modelo de los tres factores de la WISC-R se reprodujo en todos los grupos étnicos, excepto en el alumnado pápago y afroamericano, en el que únicamente se encontraron los factores Comprensión verbal y Organización perceptiva. Resultados semejantes se obtuvieron en la investigación de Zarske et al. (1981) en una muestra de alumnado, 193 navajos y 50 pápagos, con dificultades de aprendizaje; los resultados indicaron la presencia de dos factores muy similares al modelo Verbal-Perceptivo, sin embargo el factor Independencia de la distracción no surgió en estos grupos. A diferencia de los dos estudios anteriores, McShane y Plas (1982) encontraron una solución de tres factores que no presentaba semejanzas con la estructura interna propuesta en la WISC-R en una muestra de 77 niños y niñas ojibwas. Las investigaciones posteriores sobre este tema se realizaron con la Escala de Inteligencia de Wechsler para Niños-Tercera Edición (WISC-III; Wechsler, 1991) y la WISC-IV (Wechsler, 2003). Utilizaron el AFC y los resultados indicaron un buen ajuste de las estructuras factoriales de estas escalas a los datos observados (Kush & Watkins, 2007; Nakano & Watkins, 2013). Kush y Watkins (2007) analizaron la validez estructural de la WISC-III en una muestra de 344 alumnos y alumnas de educación especial de distintas etnias nativo americanas. Los resultados pusieron de manifiesto que el modelo de los cuatro factores oblicuos presentó el mejor ajuste en esta muestra comparado con otros modelos. Por otro lado, Nakano y Watkins (2013) analizaron la validez estructural de la WISC-IV en una muestra de 176 alumnos y alumnas nativo americanos de educación especial, los resultados del AFC replicaron la estructura factorial de los cuatro factores de primer orden y un factor de inteligencia general de segundo orden que explicó el 69% de la varianza común y el 33% de la total.

Las normas de *Standards for educational and psychological testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), establecen la importancia de obtener evidencia de la validez del constructo que

un instrumento pretende medir en las distintas poblaciones en las que se va a utilizar como una de las medidas para lograr una evaluación equitativa. No existe evidencia sobre la validez de constructo de la versión para México de la WISC-IV en muestras indígenas de este país. Por un lado, la validez de esta versión se determinó en una muestra de estandarización que no incluyó a los grupos indígenas (Fina et al., 2012) y, por otro lado, no existe ninguna publicación sobre este tema realizada con muestras de niños y niñas indígenas de México. Las investigaciones que han obtenido evidencia de validez de constructo a través de análisis de la estructura interna de las escalas Wechsler en nativo americanos se han realizado en Estados Unidos, con las versiones originales de estas escalas y con muestras de indígenas de ese país (Kush & Watkins, 2007; McShane & Plas, 1982; Nakano & Watkins, 2013; Reschly, 1978; Zarske et al., 1981). Esta falta de evidencia sobre la validez de constructo de la WISC-IV en muestras de indígenas de México pone en duda si esta escala está midiendo las mismas capacidades intelectuales en este alumnado que en aquel que formó parte del estudio de validación de la versión de la WISC-IV para ese país, lo cual presenta consecuencias importantes en la interpretación de las puntuaciones en estas poblaciones infantiles y en las orientaciones psicoeducativas que se pueden derivar. Por estos motivos, el objetivo del presente estudio fue analizar la validez estructural de la versión para México de la WISC-IV (Wechsler, 2007) en una muestra de alumnado indígena yaqui a través de un estudio de la matriz de intercorrelaciones de las pruebas y de un análisis factorial exploratorio y otro confirmatorio de los datos.

Basándose en los estudios anteriores (Fina et al., 2012; Wechsler, 2003), se esperan obtener en la matriz de intercorrelaciones coeficientes de correlación de bajos a moderados entre las pruebas de la escala, correlaciones más altas entre las pruebas de cada índice (sobre todo en CV) y entre pruebas de diferentes índices con saturaciones elevadas en *g*, y coeficientes de correlación muy bajos entre Claves y el resto de las pruebas.

En relación a los resultados del AFE, a partir de los resultados obtenidos por Sattler (2010)

sobre los hallazgos analítico factoriales relacionados con la edad, se espera que se reproduzca, en general, el modelo de los cuatro factores propuesto en la WISC-IV (Wechsler, 2003), pero con algunas excepciones en las cargas factoriales de las pruebas principales: Semejanzas, Conceptos con dibujos, Vocabulario, Matrices y Búsqueda de símbolos.

En cuanto a los resultados del AFC, de acuerdo a los resultados obtenidos por Fina et al. (2012) y de los estudios que han utilizado esta técnica estadística para analizar la estructura interna de las escalas Wechsler en alumnado nativo americano (Kush & Watkins, 2007; Nakano & Watkins, 2013), se conjetura que existirá un ajuste del modelo jerárquico de los cuatro factores de primer orden y un factor general de segundo orden a los datos observados.

Método

Participantes

La muestra está formada por 73 alumnos y 78 alumnas ($N=151$) de 5 escuelas públicas situadas en las comunidades yaquis de 2° ($n=121$) y 3° ($n=30$) de primaria. La media de edad del alumnado de segundo es de 7.6 ($SD=.45$) y de los de tercero de 8.6 ($SD=.27$). El 15% estaba recibiendo atención psicológica de la Unidad de Servicios de Apoyo a la Educación Regular (USAER) por problemas de aprendizaje y/o emocionales. Todo el alumnado hablaba español, el 26.5% hablaba también el idioma yaqui, el 13.2% lo hablaba un poco y el resto no lo hablaba; ninguno estaba en tratamiento médico o psiquiátrico ni tenía problemas visuales y/o auditivos. El tipo de muestra es incidental, el personal de 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, 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, 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 los indígenas en edad escolar (6-14 años) no asisten a la escuela y casi la mitad de la población indígena mencionaron como último año de estudios la primaria (Centro de Estudios Sociales y de Opinión Pública, 2011). En la comunidad yaqui el 15% de la población de 15 años o más no alcanzó la educación básica, solo el 12% cuenta con la educación media superior y el 5% superior (Instituto Nacional de Estadística y Geografía, 2010).

Instrumento

La WISC-IV (Wechsler, 2007) es un instrumento clínico de aplicación individual para evaluar la inteligencia de personas de 6 a 16 años y 11 meses de edad. Consta de 15 pruebas (10 principales y 5 optativas) que se organizan en cuatro puntuaciones que representan el funcionamiento intelectual de dominios cognoscitivos específicos (CV, RP, MT y VP), así como una puntuación compuesta que representa la capacidad intelectual general del niño (CI Total).

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 al profesorado, alumnado y padres de distintos centros escolares de la zona el objetivo de la investigación, en qué consistiría la participación del alumnado 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 de los padres y asentimiento del alumnado).

Tres psicólogas aplicaron la WISC-IV al alumnado de los grupos que aceptó participar en la investigación en instalaciones de los centros escolares, dentro del horario escolar y durante el ciclo escolar 2015-2016. Se administraron únicamente las pruebas principales según las instrucciones del manual, en una o dos sesiones, dependiendo del ritmo de ejecución del niño o de la niña. Se utilizaron los baremos mexicanos de la WISC-IV (Wechsler, 2007) para la transformación de las puntuaciones directas a típicas.

Se obtuvo el acuerdo intercalificadores únicamente en Semejanzas, Vocabulario y Comprensión cuya corrección requiere un juicio subjetivo, utilizando las puntuaciones directas. Se seleccionaron al azar 30 protocolos de la WISC-IV que fueron calificados independientemente por dos psicólogas. Los resultados mostraron 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 de acuerdo intercalificadoras, el resto de los protocolos se dividió entre las dos evaluadoras y fue corregido una sola vez. Las puntuaciones se ingresaron en el programa IBM SPSS Statistics 21 y se sometieron a un procedimiento de doble comprobación.

Análisis estadístico

El AFE se realizó con el programa IBM SPSS Statistics 21. En la extracción de los factores se usó el método de ejes principales y la rotación oblicua (Promax), tal como se llevó a cabo en la estandarización estadounidense de la WISC-IV (Wechsler, 2003). El AFC se realizó con el programa IBM SPSS Amos 23 (Arbuckle, 2016) y

se utilizó el método de estimación de máxima verosimilitud, después de determinar que el tamaño de la muestra era suficiente y analizar la normalidad univariada y multivariada de la distribución de los datos. En relación al tamaño de la muestra, se considera que entre 100 y 200 participantes son suficientes para obtener una solución factorial estable en modelos no muy complejos (Tabachnick & Fidell, 2001). Arias (2008) menciona que el número mínimo de observaciones disponibles debería ser en torno a 150 o, al menos, 5 observaciones por cada parámetro a estimar. En el presente estudio los parámetros a estimar fueron 28, por lo tanto se considera suficiente la muestra de 151 estudiantes.

El modelo que se sometió a contrastación corresponde al jerárquico con un factor de segundo orden (factor general de inteligencia) y cuatro factores de primer orden (CV, RP, MT y VP; Keith et al., 2006). En la identificación del modelo se impusieron restricciones al número de parámetros a estimar, en concreto se fijó el valor "1" a la saturación de la primera variable observada de cada factor de primer orden, a la saturación entre el factor de segundo orden (g) y el primer factor de primer orden (CV) y a las cargas entre las varianzas únicas o errores y las variables observadas.

Para evaluar el ajuste del modelo a los datos se calcularon medidas de ajuste absoluto (la prueba Chi-cuadrado, el Índice de Bondad de Ajuste [Goodness of Fit Index, GFI], el Error de Aproximación Cuadrático Medio [Root Mean Square Error of Approximation, RMSEA] y su índice asociado PClose) y medidas de ajuste incremental (el Índice de Ajuste Comparado [Comparative Fit Index, CFI] y el Índice Ajustado de Bondad de Ajuste [Adjusted Goodness of Fit Index, AGFI]).

Resultados

En la Tabla 1 se presentan las medias y desviaciones estándar de las puntuaciones escalares de las pruebas y de los índices de la WISC-IV obtenidas en la muestra. Las puntuaciones escalares medias de Semejanzas, Sucesión de números y letras, y Comprensión fueron inferiores al promedio (menores a una desviación estándar de la media 10). En el resto de

las pruebas las puntuaciones medias se situaron en el promedio. Según el sistema descriptivo normativo (Franagan & Kaufman, 2004), las puntuaciones medias de los índices RP y VP se situaron en la categoría promedio (de 85 a 115 puntuación típicas), mientras que las de los índices CV y MT en la de promedio bajo (de 70 a 84 puntuación típicas). La puntuación media del CIT se clasificó en la categoría promedio-bajo. La ejecución más baja correspondió al índice CV y la más alta a VP.

El análisis de la distribución de las puntuaciones de las pruebas indicó que los valores de asimetría y curtosis no fueron superiores a uno en valor absoluto, a excepción de Matrices que fueron 1.67 en asimetría y 6.50 en curtosis. Según West, Finch y Curran (1995) cuando los valores no sean superiores a 2 en la asimetría y a 7 en curtosis las puntuaciones se distribuyen con una moderada anormalidad univariada y, en estos casos, el método de estimación de máxima verosimilitud suele funcionar adecuadamente. En relación al análisis de la normalidad multivariada, el coeficiente de Curtosis Multivariada de Mardia obtenido a través del programa Amos 23 fue de 2.378, inferior a 10, que es el valor a partir del cual se considera que existen una desviación importante de la normalidad multivariada (The University of Texas at Austin, 16 de agosto de 2017).

Tabla 1. Estadísticos descriptivos en la WISC-IV de la muestra (N=151)

Pruebas e Índices	M	DE	Asimetría	Curtosis
Diseño con cubos	7.7	2.8	.62	.96
Semejanzas	5.3	2.4	-.09	-.44
Retención de dígitos	7.2	2.7	-.34	-.10
Conceptos con dibujos	8.1	2.7	-.36	.19
Claves	8.1	3.0	.57	.38
Vocabulario	7.2	2.8	.51	.68
Sucesión de números y letras	5.9	3.2	.28	-.47
Matrices	7.2	1.7	1.67	6.50
Comprensión	6.3	3.4	.03	-.84
Búsqueda de símbolos	7.9	2.5	-.38	.41
Índice CV	78.4	12.9	.13	-.38
Índice RP	85.8	11.0	.54	1.59
Índice MT	80.1	14.2	-.15	-.41
Índice VP	89.0	12.5	.32	.25
CIT	78.8	11.4	.21	.26

Estudio de la matriz de intercorrelaciones de las pruebas

En la Tabla 2 se presenta la matriz de intercorrelaciones de las puntuaciones de las pruebas. Todos los coeficientes de correlación fueron superiores a cero y, en general, la mayoría fueron significativos, oscilaron de .56 a .18 con un predominio de coeficientes bajos, a excepción de las correlaciones de Claves con: Diseño con cubos, Retención de dígitos, Conceptos con dibujos, Matrices y Comprensión que fueron no significativas. Las correlaciones entre las pruebas de CV fueron las más elevadas, concretamente entre Vocabulario y Comprensión (.56), y Vocabulario y Semejanzas (.41); las pruebas de este índice presentaron coeficientes de correlación bajos con las de RP, con las de MT se situaron entre bajos y moderados, y con las de VP bajos. Respecto al índice RP, los coeficientes de correlación entre sus pruebas fueron bajos; Diseño con cubos presentó una correlación con Conceptos con dibujos de .26 y con Matrices de .31, y entre estas dos últimas pruebas el coeficiente de correlación también fue de .31; las pruebas de este índice presentaron coeficientes de correlación bajos con las de otros índices. Las pruebas que conforman MT obtuvieron una correlación moderada entre sí (.44), sin embargo Retención de dígitos presentó también una correlación moderada con Vocabulario y correlaciones bajas con el resto de las pruebas de otros índices, Sucesión de números y letras obtuvo correlaciones bajas con las pruebas de otros índices. Con respecto a las pruebas de VP, entre Claves y Búsqueda de Símbolos se obtuvo un coeficiente de correlación bajo (.23), Búsqueda de Símbolos también presentó correlaciones bajas con las pruebas de otros índices y Claves entre bajas y muy bajas.

Análisis Factorial Exploratorio

El análisis de la adecuación de la muestra indicó que la matriz de correlaciones fue apropiada para el análisis factorial exploratorio. El resultado de la prueba de Esfericidad de Bartlett fue estadísticamente significativo ($\chi^2=322.244$, $gl=45$, $p<.000$); el valor de la prueba de adecuación muestral de Kaiser-Meyer-Olkin fue .832, superior a .500; y el determinante de la matriz fue .110, cercano a 0.

Tabla 2. Intercorrelaciones de las puntuaciones de las pruebas de la WISC-IV

Pruebas	DC	SE	RD	CD	CL	VB	NL	MT	CM	BS
Diseño con cubos (DC)										
Semejanzas (SE)	.18*									
Retención de dígitos (RD)	.33**	.34**								
Conceptos con dibujos (CD)	.26**	.30**	.34**							
Claves (CL)	.14	.32**	.15	.13						
Vocabulario (VB)	.25**	.41**	.46**	.30**	.26**					
Sucesión de números y letras (NL)	.30**	.31**	.44**	.25**	.21**	.33**				
Matrices (MT)	.31**	.28**	.25**	.31**	.12	.31**	.33**			
Comprensión (CM)	.22**	.26**	.32**	.30**	.01	.56**	.21**	.22**		
Búsqueda de símbolos (BS)	.24**	.27**	.30**	.19*	.23**	.33**	.34**	.25**	.28**	

** $p < .01$, * $p < .05$

Tabla 3. Saturación en g y cargas del patrón factorial exploratorio en las pruebas de la WISC-IV

Pruebas	Saturación en g		Patrón factorial			
	Primer factor no rotado	% Varianza	CV	MT	RP	VP
Semejanzas	.624	38.9	.117	.023	.204	.393
Vocabulario	.736	54.1	.693	.062	-.061	.189
Comprensión	.591	34.9	.872	-.063	.035	-.196
Diseño con cubos	.539	29.0	-.037	.355	.272	-.073
Conceptos con dibujos	.573	32.8	.033	-.068	.644	.021
Matrices	.569	32.3	-.049	.252	.385	-.012
Retención de dígitos	.691	47.7	.185	.455	.103	-.032
Sucesión de números y letras	.638	40.7	-.094	.809	-.048	-.014
Claves	.390	15.21	-.126	-.023	-.030	.738
Búsqueda de símbolos	.576	33.1	.154	.368	-.057	.123

En la Tabla 3 se presentan las saturaciones del primer factor no rotado que proporcionan información acerca de g o inteligencia general (Sattler, 2010). Este método asume que g influye indirectamente en las pruebas mediante su relación con los cuatro factores extraídos (Flanagan & Kaufman, 2004). Vocabulario es una buena medida de g , el resto de las pruebas se consideran medidas adecuadas, salvo Claves que fue una medida deficiente. La cantidad de la varianza común y total explicada por g fue de 76.8% y 35.8%, respectivamente.

Los resultados de la rotación oblicua de los factores mostró el siguiente orden de los factores: CV, MT, RP y VP. Del total de las pruebas, siete obtuvieron cargas factoriales más altas (superiores a .30) en los factores que les corresponden de acuerdo al modelo de los cuatro factores de la WISC-IV (Wechsler, 2003) y tres (Semejanzas, Diseño con cubos y Búsqueda de símbolos) presentaron saturaciones más altas en otros factores del modelo.

En el factor CV, Vocabulario y Comprensión presentaron cargas superiores a .30, sin embargo Semejanzas presentó la saturación más alta en el factor VP. Las pruebas que conforman el factor MT presentaron saturaciones superiores a .30 en este factor, sin embargo Diseño con cubos y Búsqueda de Símbolos, que corresponden a otros factores, presentaron las cargas factoriales más altas en MT. En el factor RP, Conceptos con dibujos y Matrices presentaron cargas factoriales superiores a .30. Diseño con cubos, en cambio, presentó la saturación más alta en el factor MT. De las dos pruebas que integran el factor VP, únicamente Claves presentó la carga factorial más alta en este factor. Búsqueda de símbolos presentó su carga factorial más alta en MT.

Análisis Factorial Confirmatorio

En la Figura 1 se presenta la solución del AFC del modelo jerárquico de los cuatro factores de primer orden y un factor general de inteligencia (g) de segundo orden de la WISC-IV en la muestra.

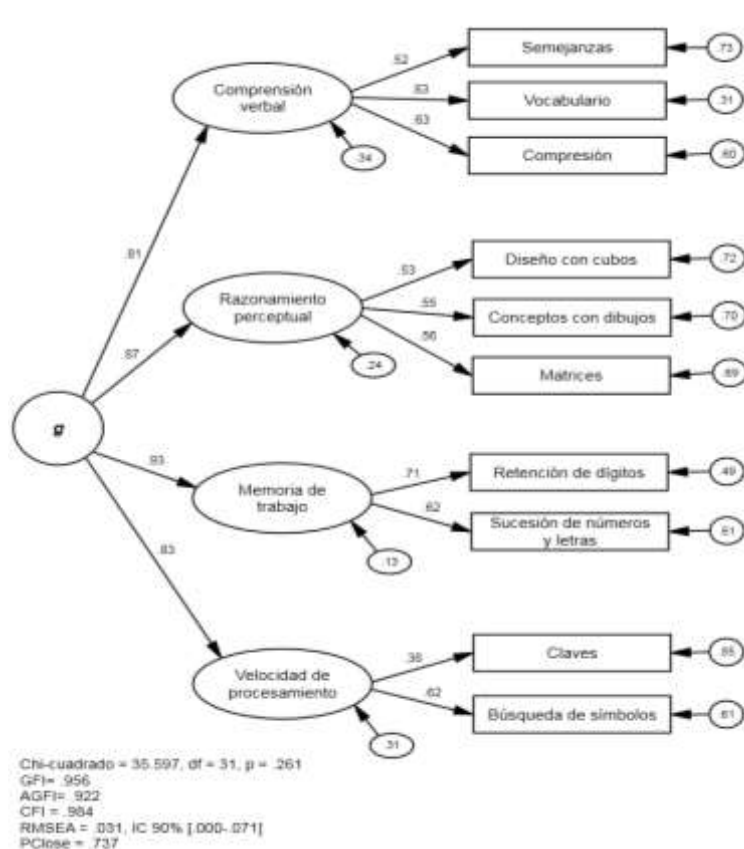


Figura 1. Solución factorial del modelo jerárquico de segundo orden de la WISC-IV en la muestra de alumnado. Coeficientes de regresión estandarizados e índices de bondad de ajuste del modelo

Los índices de bondad de ajuste mostraron un ajuste perfecto del modelo a los datos observados (Arias, 2008): La prueba Chi-cuadrado fue no significativa, los valores GFI y AGFI fueron superiores a .90, el valor CFI superior a .95, RMSEA fue inferior a .05 y el valor PClose (.737) permitió aceptar la hipótesis nula de que la población RMSEA no es mayor a .05.

Como se aprecia en la Figura 1 los coeficientes de regresión estandarizados en el factor de segundo orden fueron altos y estadísticamente significativos ($\beta = .81-.93$, $p < .005$). Los factores que presentaron mayor carga factorial en g fueron MT y RP, seguidos de VP y, por último, CV. Por otro lado, los coeficientes de regresión entre los factores de primer orden y las pruebas fueron significativos y, en general, moderados (el rango osciló entre .52 y .83, $p < .005$) a excepción de Claves con un valor bajo (.38).

Discusión

Los resultados de la investigación proporcionan evidencia de validez estructural de

la WISC-IV (2007) en la muestra de alumnado yaqui. Se aceptan las hipótesis planteadas sobre el estudio de la matriz de intercorrelaciones y el AFC, sin embargo los resultados del AFE mostraron algunas diferencias con la hipótesis planteada acerca de este análisis. Por otro lado, se constató que la WISC-IV es una medida adecuada de la inteligencia general (g) en esta muestra puesto que explica la mayor cantidad de varianza común (76.8%) y total (35.8%). Este resultado es consistente con los de las investigaciones anteriores sobre esta escala (Nakano & Watkins, 2013; Watkins, 2006).

Con respecto al análisis de la matriz de intercorrelaciones, se obtuvieron coeficientes de correlación de bajos a moderados, con un predominio de los bajos, entre las pruebas. Lo cual confirma que todas miden un factor de inteligencia general (g). Las pruebas de un determinado índice correlacionaron, en general, con más intensidad entre ellas que con las de otros índices, sobre todo las que integran el índice CV, pero también se obtuvieron coeficientes de correlación semejantes entre pruebas de diferentes

índices en la medida en que comparten determinadas capacidades y presentan saturaciones altas en *g*. Por ejemplo, las pruebas de MT correlacionaron moderadamente entre sí, pero a la vez presentaron correlaciones moderadas con pruebas de CV posiblemente por las demandas de comprensión auditiva que comparten y porque sus cargas en *g* fueron de las más elevadas de la escala. Por otro lado, se encontraron coeficientes de correlación muy bajos y algunos no significativos entre Claves y el resto de las pruebas.

Este patrón de intercorrelaciones es muy semejante al encontrado en las matrices de los grupos de 7 y 8 años de edad de la estandarización estadounidense (Wechsler, 2003) y de la muestra de estandarización para México de la WISC-IV (Fina et al., 2012). En estas dos investigaciones la mayoría de los coeficientes de correlación entre las pruebas se situaron entre moderados y bajos, con un predominio de los bajos en las matrices de los grupos de 7 y 8 años de edad de la muestra de estandarización estadounidense formada por 200 niños y niñas en cada edad y un predominio de los moderados en la matriz de la muestra de estandarización mexicana, probablemente por el mayor tamaño de la muestra. Las correlaciones más altas de estas matrices también fueron entre las pruebas del índice CV y las más bajas (de las pruebas principales) correspondieron a Claves con otras pruebas. La diferencia más destacada de esta comparación consistió en que en el presente estudio MT fue el índice que presentó correlaciones más altas entre las pruebas que lo componen después de las del índice CV, en cambio en las matrices de intercorrelación de las muestras de estandarización estadounidense y mexicana el índice RP fue el que presentó las intercorrelaciones más altas después del de CV. Esta diferencia está relacionada con las saturaciones *g* de las pruebas, es decir, el grado en el que miden la inteligencia general. En el presente estudio las pruebas que integran MT obtuvieron saturaciones *g* más elevadas que las pruebas de RP, en cambio los resultados sobre *g* obtenidos en la estandarización estadounidense y mexicana indicaron saturaciones más altas en las pruebas de RP, sobre todo Diseño con cubos y Matrices, que las pruebas que componen el índice MT.

Los resultados del AFE no muestran una reproducción totalmente semejante al modelo de los cuatro factores de la WISC-IV (Wechsler, 2003). Existen algunas diferencias que tienen que ver con inconsistencias en las cargas factoriales de determinadas pruebas y el orden de extracción de los factores. La primera diferencia se tuvo en cuenta en la hipótesis planteada sobre los resultados del AFE, según la cual se esperaba encontrar una reproducción del modelo de los cuatro factores de la escala, pero con algunas excepciones en las cargas factoriales de determinadas pruebas. Las excepciones que se han encontrado corresponden a Semejanzas, Búsqueda de símbolos y Diseño con cubos, que presentaron cargas factoriales más altas en otros factores del modelo. La investigación realizada por Sattler (2010) mostró que Semejanzas y Búsqueda de símbolos también presentaron cargas factoriales inconsistentes con el modelo de los cuatro factores de la WISC-IV, sin embargo Diseño con cubos presentó siempre su carga factorial más alta en el factor RP en todos los grupos de edad. La variación estructural en relación al modelo de los cuatro factores de la WISC-IV en la muestra de alumnado yaqui es consistente con los resultados de los estudios anteriores que muestran un patrón diverso relacionado con la edad (Sattler, 2010; Wechsler, 2003). Sattler (2010) argumenta que las razones de esta variación no están claras, pueden deberse al hecho de que todas las pruebas tienen una relación diferencial con *g*, por errores de medida o tendencias del desarrollo. En relación a la segunda diferencia, el orden de los factores extraído en el AFE no coincide con el del modelo de los cuatro factores, los factores MT y RP mostraron un orden inverso, el factor MT ocupó el segundo lugar y el de RP el tercer lugar, posiblemente debido a la mayor saturación en *g* que presentan en esta muestra las pruebas del factor MT respecto a las de RP.

Los resultados del AFC respaldan el modelo jerárquico de los cuatro factores de primer orden y un factor general (*g*) de segundo orden de la WISC-IV en la muestra. Los índices de bondad de ajuste que se calcularon indicaron un ajuste perfecto de este modelo a los datos observados. Los coeficientes de regresión estandarizados del factor de segundo orden (*g*) fueron altos y estadísticamente significativos, los factores que

obtuvieron la mayor carga factorial en *g* fueron, en primer lugar, MT y, en segundo lugar, RP. Por otro lado, los coeficientes de regresión de los factores de primer orden (índices) hacia las pruebas fueron significativos y, en promedio, se situaron en un rango moderado; el más alto fue el de Vocabulario (.83) y el más bajo Claves, que con un valor de .38 indica una escasa fiabilidad ($R^2=.14$) y que no es una buena medida del factor VP. Probablemente, la aplicación de las dos formas diferentes de esta prueba según la edad (Claves A para el alumnado de 6-7 años y Claves B para el de 8 años de edad) influyeron en un aumento de su varianza de error. La comparación de estos resultados con los obtenidos en la muestra de estandarización de la versión mexicana de la WISC-IV (Fina et al., 2012) y con las investigaciones que han utilizado el AFC para analizar la configuración interna de las escalas Wechsler en alumnado nativo americano (Kush & Watkins, 2007; Nakano & Watkins, 2013) muestra coincidencias. Fina et al. (2012) también encontraron un buen ajuste del modelo jerárquico de la WISC-IV a los datos observados y, al igual que en el presente estudio, MT y RP presentaron las cargas más altas del segundo factor (*g*), Vocabulario obtuvo la carga más alta de los factores de primer orden y Claves la más baja – según estos autores posiblemente por su baja fiabilidad-. La única investigación que ha analizado la validez estructural de la WISC-IV utilizando la técnica del AFC en muestras de alumnado nativo americano corresponde a la de Nakano y Watkins (2013) cuyos resultados presentan semejanzas con los de la presente investigación. Estos autores encontraron que el modelo que presentaba el mejor ajuste a los datos observados fue el jerárquico, que la carga más alta del factor de segundo orden también la obtuvo MT y que, de los factores de primer orden la carga más alta la presentó Vocabulario, sin embargo la más baja no correspondió a Claves sino a Búsqueda de Símbolos.

Las discrepancias entre los resultados obtenidos en el AFE y el AFC pueden atribuirse lógicamente a que los modelos estructurales que se analizaron son distintos, pero también a diferencias entre estas técnicas estadísticas. En el AFE se estudió el modelo de los cuatro factores, en cambio en el AFC se sometió a contrastación

una estructura jerárquica con un factor de segundo orden (*g*) y cuatro de primer orden. Respecto a las diferencias entre las técnicas estadísticas, una de las características del AFE es la indeterminación factorial, es decir, la estructura factorial que se aporta es una solución plausible dentro del conjunto infinito de soluciones posibles (Gorsuch, 1983). Asimismo, como este tipo de análisis no se realiza sobre una base teórica de las relaciones entre los datos, todas las variables cargan en todos los factores, a diferencia del AFC que busca probar un modelo y sólo las variables evaluadas en cada factor presentan cargas factoriales en esos factores (Fernández, 2015).

Se obtuvo un bajo rendimiento del alumnado yaquí en la WISC-IV según las normas de la muestra de estandarización para México (Wechsler, 2007). Todas las puntuaciones medias de los índices y el CIT se situaron por debajo de la media normativa. Este patrón de puntuaciones es semejante a los obtenidos por Nakano y Watkins (2013) y Kush y Watkins (2007) en muestras de alumnado nativo americano con necesidades de educación especial y las investigaciones que han encontrado que los grupos pertenecientes a niveles socioeconómicos desfavorecidos presentan puntuaciones más bajas en los tests de capacidad intelectual (Coronel, Lacunza, & Contini, 2006). Si bien antes se pensaba que las diferencias entre las puntuaciones medias de personas procedentes de diferentes grupos étnicos eran un indicador de sesgo cultural de los tests, en la actualidad estas diferencias no se consideran en sí mismas un criterio para detectar este aspecto (Sattler, 2010). Wechsler (1971) planteó que estas diferencias eran un indicador de las desigualdades de la sociedad y de cómo las diferentes oportunidades médicas, políticas, económicas y sociales tenían impacto en las habilidades intelectuales de las personas. De acuerdo con esta explicación, consideramos que las diferencias de ejecución observadas en la presente investigación son consecuencia de las menores oportunidades educativas y socioculturales que presenta los niños y niñas de este grupo indígena respecto al alumnado mexicano que formó parte de la muestra normativa de esta escala.

En conjunto, los resultados obtenidos aportan evidencia empírica de validez estructural de la versión para México de la WISC-IV en la muestra

de alumnado yaqui. Se puede concluir que esta escala no presenta sesgo cultural en lo que respecta a este tipo de validez. La versión mexicana de la WISC-IV parece ser un instrumento con utilidad clínica para evaluar la capacidad intelectual de los niños y niñas de 7 y 8 años de edad de esta población, que proporciona una buena medida de la inteligencia general. Sin embargo, teniendo en cuenta el bajo rendimiento obtenido en esta población infantil y el hecho de que en la estandarización de esta escala en México no se incluyeron grupos indígenas, se recomienda que las interpretaciones de las puntuaciones se realicen en el marco de unas prácticas de evaluación psicológica que integren los resultados de esta escala con otros métodos y fuentes, los cuales tengan en cuenta las acciones de las personas en sus hábitats cotidianos (Casullo, 2009) para conseguir, en la medida de lo posible, diagnósticos y orientaciones adecuados.

Entre las limitaciones de esta investigación destacamos, en primer lugar, que únicamente se utilizaron las pruebas principales para probar un modelo jerárquico donde algunos factores de primer orden estaban formados por sólo dos variables manifiestas, lo cual disminuye la probabilidad de que los constructos sean adecuadamente representados para formar un factor coherente. Según Velicer y Fava (1998) se requieren como mínimo tres variables por factor para reproducir un patrón factorial. En segundo lugar, el tamaño de la muestra y el tipo de muestreo utilizado limitan la generalización de los resultados a otras edades y grupos indígenas de México. Son necesarias futuras investigaciones con muestras más amplias que garanticen la estabilidad de los resultados obtenidos en los análisis factoriales y que abarquen otros grupos de edad para obtener información sobre la validez estructural de la escala en función de esta variable. Finalmente, sería conveniente que las siguientes adaptaciones de esta escala en México incluyeran en la muestra de estandarización los grupos indígenas de este país para incrementar la validez de las interpretaciones de las puntuaciones que obtienen los niños y niñas de estos grupos étnicos.

Referencias

- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Autor.
- Arbuckle, J. (2016). *Amos™ 24 user's guide*. Chicago, IL: IBM, SPSS.
- Arias, B. (2008). Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS, y SAS. En M. A. Verdugo, M. Crespo, M. Badía y B. Arias (Coord.), *Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales* (pp. 75-120). VI Seminario Científico SAID. Salamanca: Publicaciones del INICO.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. doi:<http://dx.doi.org/10.1037/h0046016>
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. En D. P. Flanagan, U. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122-130). New York: Guilford.
- Casullo, M. M. (2009). La evaluación psicológica: Modelos, técnicas y contextos. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 27(1), 9-28.
- Centro de Estudios Sociales y de Opinión Pública (2011). *Situación de los indígenas, n° 16*. Recuperado de <http://www3.diputados.gob.mx/camara/contenido/view/full/75718>
- Comisión Nacional para el Desarrollo de los Pueblos Indígenas (2017). *Indicadores socioeconómicos de los pueblos indígenas de México, 2015*. Recuperado de <https://www.gob.mx/cdi/documentos/indicadores-socioeconomicos-de-los-pueblos-indigenas-de-mexico-2015>
- Consejo Nacional de Población. (2010). *Índice de marginación por localidad 2010*. Recuperado de <http://www.conapo.gob.mx>

- Coronel, C. P., Lacunza, A. B., & Contini, N. (2006). Las Habilidades Cognitivas en niños privados culturalmente. Resultados preliminares de la primera fase de evaluación. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 22(2), 49-74.
- Fernández, A. (2015). Aplicación del análisis factorial confirmatorio a un modelo de medición del rendimiento académico en lectura. *Ciencias Económicas*, 33(2), 39-66. doi:<http://dx.doi.org/10.15517/rce.v33i2.22216>
- Fina, A. D., Sánchez-Escobedo, P., & Hollingworth, L. (2012). Annotations on Mexico's WISC-IV: A validity study. *Applied Neuropsychology: Child*, 1, 6-17. doi:10.1080/21622965.2012.665771
- Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV Assessment*. Hoboken, NJ: John Wiley & Sons.
- Gorsuch, R. L. (1983). *Factor analysis* (2ª ed.). Hillsdale, NJ: Erlbaum.
- Instituto Nacional de Estadística y Geografía (2010). *Censo de población y vivienda 2010*. Recuperado de <http://www3.inegi.org.mx/rnm/index.php/catalog/71>
- Keith, T. Z., Goldenring, F., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review*, 35, 108-127.
- Kush, J. C., & Watkins, M. W. (2007). Structural validity of the WISC-III for a national sample of Native American students. *Canadian Journal of School Psychology*, 22, 235-248. doi:10.1177/0829573507303056
- McShane, D. A., & Plas, J. M. (1982). WISC-R factor structures for Ojibwa Indian children. *White Cloud Journal of American Indian Mental Health*, 2(4), 18-22.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational measurement: Issues and Practice*, 14, 5-8.
- Nakano, S., & Watkins, M. W. (2013). Factor structure of the Wechsler Intelligence Scales for Children-Fourth Edition among referred Native American students. *Psychology in the Schools*, 50(10), 957-968. doi:10.1002/pits.21724
- Prifitera, A., Saklofske, D. H., Weiss, L. G., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. En A. Prifitera, D. H. Saklofske, & L. G. Weiss, *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 3-32). London, UK: Elsevier.
- Reschly, D. J. (1978). WISC-R factor structures among Anglos, Blacks, Chicanos, and Native-American Papagos. *Journal of Consulting and Clinical Psychology*, 46(3), 417-422. doi:<http://dx.doi.org/10.1037/0022-006X.46.3.417>
- Sattler, J. M. (2010). *Evaluación infantil. Fundamentos cognitivos* (5ª ed.). México D. F.: El Manual Moderno.
- Secretaría de Educación Pública (2012). *Educación pertinente e inclusiva. La discapacidad en educación indígena. Guía-Cuaderno 3: Atención educativa de alumnos y alumnas con discapacidad intelectual*. México D.F.: Autor.
- Secretaría de Educación Pública (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*. Hermosillo, México: Autor.
- Spicer, E. (1994). *Los Yaquis: Historia de una cultura*. México D. F.: Universidad Autónoma de México.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4ª ed.). Boston, MA: Allyn and Bacon.
- The University of Texas at Austin. (16 de agosto de 2017). *Software FAQs. AMOS* [página web]. Recuperado de <https://stat.utexas.edu/software-faqs/amos>
- Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessments issues*. Thousand Oaks, CA: Sage.
- Velicer, W. E., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3(2), 231-251. doi:10.1037//1082-989X.3.2.231
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale

- for Children-Fourth Edition. *Psychological Assessment*, 18, 123-125.
doi:10.1037/1040-3590.18.1.123
- Wechsler, D. (1971). Intelligence: Definition, theory, and the IQ. En R. Cancro (Ed.), *Intelligence: Genetic and environmental influences* (pp. 50-55). New York: Gruene and Stratton.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2007). *WISC-IV: Escala Wechsler de Inteligencia para Niños-IV: Manual técnico*. México D.F.: El Manual Moderno.
- West, S. G., Finch, J. P., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.
- Zarske, J. A., Moore, C. L., & Peterson, J. D. (1981). WISC-R factor structure for diagnosed learning disabled Navajo and Papago children. *Psychology in the Schools*, 18, 402-407.
doi:10.1002/1520-6807(198110)18:4<402::AID-PITS2310180404>3.0.CO;2-J