

CALIDAD DE LA ADAPTACIÓN Y TRADUCCIÓN DEL WISC-IV PARA LA POBLACIÓN COLOMBIANA

Laura Sofía Carreño Novoa¹
Universidad Nacional de Colombia

Resumen

La adaptación y traducción de tests es un proceso que se está elaborando con mayor frecuencia en muchas culturas y países. En Colombia la Escala Weschler de Inteligencia para niños-edición 4° (WISC-IV) es usado para evaluar el área cognitiva de los niños, pero este test aún no ha sido adaptado a la población colombiana, por lo que se usan las versiones mexicanas y españolas. Este trabajo intenta responder a la pregunta de si usar adaptaciones extranjeras es bueno para la población colombiana, y de ser así cuál de las dos versiones es más apta. Para esto se hizo una comparación de medias de muestras relacionadas, y un análisis de la equivalencia estructural de las dos adaptaciones por medio del escalamiento multidimensional pesado WMDS.

Palabras Clave: adaptación, traducción, test, WISC-IV, directrices, equivalencia.

Abstract

Adaptation and translation of tests is a process that is being developed more frequently in many cultures and countries. In Colombia the Weschler Intelligence Scale for Children-edition 4th (WISC-IV) is used to evaluate the cognitive area of children, but this test has not yet been adapted to the Colombian population, so that Mexican and Spanish versions are used. This work tries to answer the question of whether to use foreign adaptations is good for the Colombian population, and if so which of the two versions is more suitable. In order to find the answer, a comparison of means of related samples was made, and an analysis of structural equivalence of the two adaptations through multidimensional scaling WMDS.

Key Words: Adaptation, Translation, WISC-IV, guidelines, equivalence.

Adaptación De Tests Psicológicos

¹ Estudiante de psicología, Universidad Nacional de Colombia
lscarrenon@unal.edu.co

La adaptación o traducción de test psicológicos es un proceso de gran relevancia en la actualidad, ya que gracias a este no solo hay un intercambio cultural, sino que también permite que otros países y culturas puedan utilizar instrumentos que tienen altos índices de confiabilidad y suficientes estudios de validez, pero en su propia lengua, es decir, que el proceso de adaptación permite que un test sea igualmente válido y confiable en una población como lo ha sido en su cultura origen. Por otra parte, la realización de estos estudios transculturales permiten comparar resultados entre esas diferentes culturas y con ellos establecer pautas comunes entre las diferentes poblaciones a las que se quiere adaptar un test (Gómez-Benito, Hidalgo-Montesinos, 2003). En los últimos años, este proceso ha incrementado tanto que llega a diferentes campos evaluativos (Muñiz y Hambleton, 1996), lo cual se ha visto reflejado en los instrumentos empleados en áreas educativa, clínica, jurídica, social, organizacional, entre otras, permitiendo un aumento multicultural y multilingüe (Muñiz y cols. 2013).

Dado que hay una gran demanda por la adaptación de test entre culturas, se hace necesario señalar los pasos a seguir, los errores comunes y los problemas presentados en el proceso de adaptación (Hambleton, 1996).

El proceso de adaptación y traducción de un test va más allá de la traducción como tal, ya que requiere que los resultados en las puntuaciones del test traducido se equiparen con el original, (Gómez-Benito, Hidalgo-Montesinos, 2003) de lo contrario, el test adaptado no tendría validez. Para alcanzar esa equivalencia hay que considerar cuatro aspectos del proceso, los cuales fueron mencionados por Hambleton (1996): el contexto cultural, aspectos técnicos, administración y finalmente interpretación y documentación de las puntuaciones. Además de estos cuatro aspectos básicos, Hambleton y Patsula (1999; en Gómez-Benito, Hidalgo-Montesinos, 2003) han sistematizado el proceso de adaptación en trece pasos consecutivos, los cuales son: 1) equivalencia de los constructos en los grupos culturales de interés, 2) decidir si se adapta un test ya existente o se desarrolla uno nuevo, 3) seleccionar traductores profesionales cualificados, 4) combinar diseños de traducción, 5) revisar la versión adaptada del test y realizar las correcciones necesarias, 6) hacer una prueba piloto con el test adaptado, 7) hacer una prueba empírica del test adaptado, 8) comparar las puntuaciones del test adaptado y original, 9) si existe interés en hacer comparaciones interculturales hay que asegurarse de la equivalencia de las versiones, 10) hacer estudios de validación, 11) documentar todo el proceso y hacer un manual para el test adaptado, 12) entrenar a los usuarios, 13) realizar un seguimiento del test adaptado.

Adicionalmente es importante observar las fuentes de error para maximizar la validez de la adaptación, y permitir, a su vez, conocer a profundidad la cultura objetivo.

Fuentes de error: Los errores en la adaptación de un test hacen que se invalide el instrumento y que ya no sea aplicable a la cultura objetivo, es por eso que es tan relevante identificarlas y controlarlas. Según Hambleton (1996), las fuentes de error se unifican en tres grupos, los cuales están relacionados con los estándares para los test psicológicos. El primer grupo corresponde a las diferencias culturales y de idioma, donde se observa el formato del test, la forma de aplicación, y la velocidad del mismo; dado que varían de una cultura a otra, es importante la búsqueda de la equivalencia de ambas poblaciones en función de estos aspectos para que las puntuaciones y la interpretación de las mismas no se vea alterado por costumbres y aspectos culturales. El segundo grupo tiene que ver con aspectos técnicos y métodos, donde se tiene que tener en cuenta la forma del propio test, los traductores (selección y entrenamiento), proceso de traducción, diseños racionales para la adaptación de tests, Diseños empíricos para establecer la equivalencia entre el test adaptado y el original. El tercer y último grupo es de interpretación de los resultados, el cual es importante para que la aplicación de un test se use para la población que va dirigida y no para la comparación entre países, pues esto puede generar malas interpretaciones debido a las diferencias culturales existentes. Por lo tanto, para realizar una buena interpretación de los resultados de un test deben tenerse en cuenta factores específicos de cada país, que sean ajenos al test. Entre estos factores se pueden encontrar la motivación, los currículos, los factores socio-políticos, económicos, educativos, entre otros (Hambleton, 1996).

Directrices Prácticas para la adaptación de test

Las directrices son las que guían el proceso de adaptación, traducción y desarrollo del test, y ofrecen pautas para corregir los errores más comunes. El objetivo es que el producto final del proceso de adaptación sea lo mas equivalente posible al original, en relación con la cultura, conceptos, medición y lingüística; las directrices se consideran patrones que guían al investigador en el proceso de adaptación (Muñiz, Elosua y Hambleton, 2013).

la Comisión Internacional de Tests (ITC) ha avanzado con este propósito y desde Año XXX ha formulado un conjunto de directrices prácticas para adaptación de tests (Hambleton, 1996); inicialmente eran 22 directrices, organizadas en cuatro secciones: Contexto (2), Construcción del test y adaptación (10), Aplicación (6) y Documentación/interpretación de las puntuaciones (4). Para el año 2013, la ITC ha enseñado una segunda edición para las

directrices de adaptación de tests de unas culturas a otras; esta nueva versión contempla 20 directrices organizadas en 6 grupos. En esta nueva edición se introduce el concepto de permisos legales para garantizar la autenticidad de la adaptación, la protección de derechos de autor y propiedad intelectual; como también pautas para la redacción del manual, y estudio de la relevancia del constructo (Muñiz y cols. 2013). Estas nuevas directrices se presentan en la tabla 1.

	categoría	número de directrices	aspectos analizados	
Fuente: Directrices adaptación adaptación tabla 1 Cols. 2013)	previas	5	Marco legal, Diseño, Evaluación del constructo	Tabla 1. para de pruebas del de la (Muñiz y
	desarrollo	5	Adaptación lingüística, Adaptación cultural, Estudios piloto	
	confirmación	4	Recogida de datos, Equivalencia, Fiabilidad, Validación	
	aplicación	2	Administración	
Escala Inteligencia IV (WISC-	puntuación e interpretación	2	Interpretación de las puntuaciones, Comparabilidad	Weschler de para Niños- IV)
	documentación	2	Cambios entre versiones, Uso correcto	

El WISC-IV fue diseñado para evaluar inteligencia en niños de 6 años hasta 16 años, se compone de 15 subpruebas (10 esenciales y 5 complementarias) que miden cuatro dominios de la inteligencia: Comprensión Verbal, Razonamiento Perceptual, Memoria de trabajo y Velocidad de Procesamiento; los cuales en su totalidad, brindan un perfil de inteligencia general del consultante. Entre las aplicaciones y usos de esta prueba, se encuentran identificar necesidades escolares especiales, detección de retraso mental, detección de capacidades intelectuales excepcionales, entre otros. Las puntuaciones compuestas de esta prueba van desde 40 hasta 160; entre 90 y 110 está el promedio de la población, de 80 a 90 promedio bajo, 70 a 80, capacidad límite y de 70 hacia abajo, retraso considerable muy bajo; mientras que de 110 a 120 es promedio alto y de 120 hacia arriba habilidades superiores. Los intervalos

de confianza son las puntuaciones donde se espera encontrar la cantidad de atributo real del sujeto, el cual puede utilizarse un porcentaje de confianza de 90 o de 95.

El Índice de Comprensión Verbal (ICV) mide la capacidad de razonamiento verbal, formación de conceptos y nivel de conocimientos adquiridos; el Índice de Razonamiento Perceptual (IRP) evalúa la capacidad de pensar con imágenes visuales, interpretar estímulos presentados de forma visual, formar conceptos abstractos y comunicación no verbal; El Índice de Memoria de Trabajo (IMT) evalúa la memoria a corto plazo, la atención, la capacidad numérica, de codificación y de auto monitoreo, y el procesamiento auditivo; y el Índice de Velocidad de Procesamiento (IVP), evalúa la discriminación perceptual, velocidad psicomotora, atención, concentración y capacidad numérica.

Comparando las dos versiones

En Colombia durante un proceso de evaluación cognitiva infantil y adolescente es bien conocido el uso del WISC-IV para hacer diagnóstico o buscar las soluciones más óptimas a las necesidades de cada uno de los sujetos evaluados; pero no existe una adaptación y validación de esta prueba que permita evaluar a los niños colombianos y comparar sus resultados con la población del mismo país, por lo que las versiones utilizadas son la mexicana y la española. En el Servicio de Atención Psicológica (SAP) de la Universidad Nacional de Colombia se utilizan ambas versiones para realizar procesos de evaluación cognoscitiva; entonces partiendo de la pregunta de investigación de que si esta prueba que ha sido utilizada en Colombia pero adaptada en otros países de habla hispana, tiene la calidad suficiente y necesaria para un buen uso de ella en esta cultura, poniendo de manifiesto la necesidad de hacer una comparación entre los resultados obtenidos en ambas versiones para poder saber cuál de las dos versiones sería más apta para usar en Colombia, específicamente en el SAP.

Para lograr lo propuesto, se utilizará la técnica estadística multi-variada con el fin de examinar el grado de equivalencia estructural del WISC-IV a través de diferentes grupos; esto se logrará por medio del Escalamiento Multidimensional Pesado (WMDS), esta es una técnica exploratoria que permite determinar si la estructura factorial de un grupo es similar a la de otro; se utilizan pesos para determinar la importancia de cada dimensión para cada grupo (Padilla, 2013).

El WMDS usa dos índices de ajuste para determinar la estructura dimensional de la prueba: el *stress*, que es la raíz cuadrada de la varianza de los residuales normalizados en un modelo

de regresión, las puntuaciones de este índice toman un valor mínimo de cero, entre mayor sea este valor, mayor distancia habrá y peor será el modelo, por lo tanto cuanto más bajo sea el índice, mejor será el ajuste. El otro índice utilizado es el coeficiente de correlación al cuadrado (*RSQ*) que informa acerca de la proporción de la variabilidad de los datos que es explicada por el modelo, los valores de este índice oscilan entre 0 y 1, y a diferencia del *stress*, los valores cercanos a 1 indican que el modelo es adecuado (Guerrero y Ramírez, s.f.).

Metodología

Participantes

Los participantes para este trabajo fueron consultantes del SAP entre 6 años y 2 meses hasta 16 años y 11 meses, quienes asistieron al servicio para realizar un proceso de evaluación cognoscitiva; en total fueron 76 sujetos a quienes se les aplicó el WISC-IV versión mexicana, de los cuales 26 eran mujeres y 50 eran hombres. La información de los participantes es confidencial, por lo tanto los datos se tomaron de acuerdo a la digitación de resultados realizada por cada uno de los psicólogos que trabajan en el SAP y que hagan parte del área de evaluación.

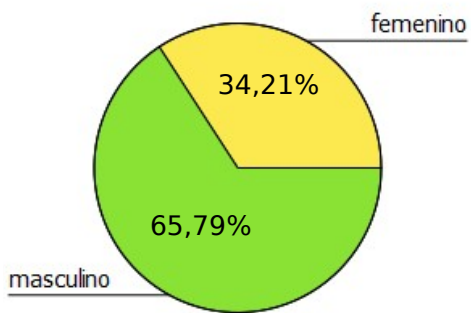
Aplicación

Como ya se mencionó, la aplicación del instrumento se realizó durante el proceso de evaluación cognoscitiva llevado a cabo para cada consultante. Pero durante este proceso sólo se aplicó el WISC-IV versión mexicana, entonces con el fin de obtener datos de la aplicación del WISC-IV versión estadounidense y asegurar la homogeneidad de la muestra y los resultados, se transformaron los baremos mexicanos a baremos estadounidenses y de esta manera se obtuvo la base de datos para ambas adaptaciones.

Resultados

Estadísticos descriptivos:

GÉNERO

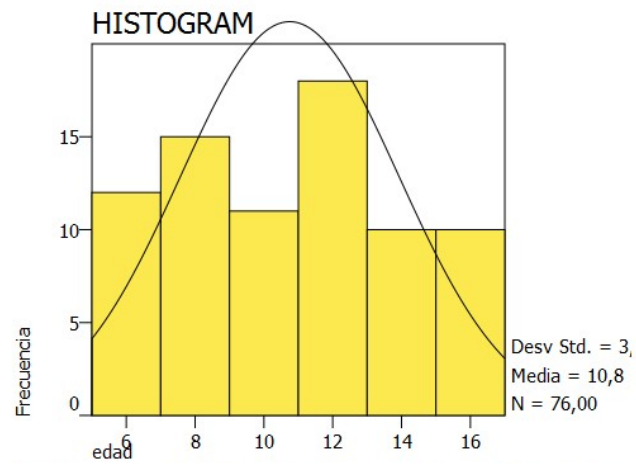


Gráfica 1

genero

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
femenino	1	26	33,77	34,21	34,21
masculino	2	50	64,94	65,79	100,00
.	.	1	1,30	Perdidos	
Total		77	100,0	100,0	

<i>N</i>	<i>Válido</i>	76
	<i>Perdidos</i>	1
<i>Media</i>		10,75
<i>Desv Std</i>		3,18
<i>Mínimo</i>		6,10
<i>Máximo</i>		16,11



Gráfica 2

Comparación de medias poblacionales.

Prueba de normalidad:

Baremos mexicanos

		ICV	IRP	IMT	IVP	CIT
N		76	76	76	76	76
Parámetros Normal	Media	85,16	89,75	86,04	80,41	82,42
	Desviación Estándar	15,55	16,12	15,86	13,55	16,29
Diferencias Más Extremas	Absoluto	,13	,11	,08	,08	,11
	Positivo	,06	,07	,08	,08	,09
	Negativo	-,13	-,11	-,08	-,06	-,11
Z de Kolmogorov-Smirnov		1,16	,92	,73	,67	,96
Sig. Asint. (2-colas)		,12	,36	,66	,76	,32

Baremos estadounidenses

		ICV_estadounidense	IRP_estadounidense	IMT_estadounidense	IVP_estadounidense	CIT_estadounidense
N		76	76	76	76	76
Parámetros Normal	Media	81,88	86,68	82,47	76,96	78,04
	Desviación Estándar	16,08	16,70	17,72	13,59	16,39
Diferencias Más Extremas	Absoluto	,12	,11	,08	,08	,07
	Positivo	,08	,06	,07	,08	,07
	Negativo	-,12	-,11	-,08	-,06	-,06
Z de Kolmogorov-Smirnov		1,09	,97	,69	,68	,64
Sig. Asint. (2-colas)		,17	,30	,72	,74	,81

De acuerdo a las tablas anteriores se observa que la distribución de los datos para los cuatro índices de la inteligencia, y la capacidad total, tanto para la aplicación de la adaptación mexicana y la estadounidense tienen una distribución normal, ya que el valor de *P* para cada una de las variables es mayor a 0,05.

Prueba t-student para muestras emparejadas.

Se realizó la comparación de cada uno de los índices del WISC-IV para la adaptación mexicana y estadounidense; para ello se hizo un análisis por medio de la prueba t-student para muestras emparejadas, dado que se trató de los mismos sujetos para cada aplicación.

A continuación se muestra la tabla de correlaciones para cada pareja, mostrando que existe una alta correlación entre cada uno de los índices para cada adaptación, pues la más baja es de 0,93 y la más alta de 0,98.

		N	Correlación	Sign.
Pareja 0	ICV_mexicano & ICV_estadounidense	76	,96	,00
Pareja 1	IRP_mexicano & IRP_estadounidense	76	,98	,00
Pareja 2	IMT_mexicano & IMT_estadounidense	76	,96	,00
Pareja 3	IVP_mexicano & IVP_estadounidense	76	,93	,00
Pareja 4	CIT_mexicano & CIT_estadounidense	76	,96	,00

A continuación está la tabla de las diferencias emparejadas, utilizando un 95% de confianza, se puede ver que existen diferencias significativas entre las puntuaciones de cada uno de los índices para cada adaptación.

	Diferencias emparejadas					t	df	Sign. (2- colas)
	Media	Desviación Estándar	Error Est. Media	Intervalo de confianza 95% de la Diferencia				
				Inferior	Superior			
Pareja 0 ICV_mexicano - ICV_estadounidense	3,28	4,44	,51	2,26	4,29	6,44	75	,00
Pareja 1 IRP_mexicano - IRP_estadounidense	3,07	3,31	,38	2,31	3,82	8,08	75	,00
Pareja 2 IMT_mexicano - IMT_estadounidense	3,57	4,92	,56	2,44	4,69	6,32	75	,00
Pareja 3 IVP_mexicano - IVP_estadounidense	3,45	5,14	,59	2,27	4,62	5,85	75	,00
Pareja 4 CIT_mexicano - CIT_estadounidense	4,38	4,48	,51	3,36	5,41	8,52	75	,00

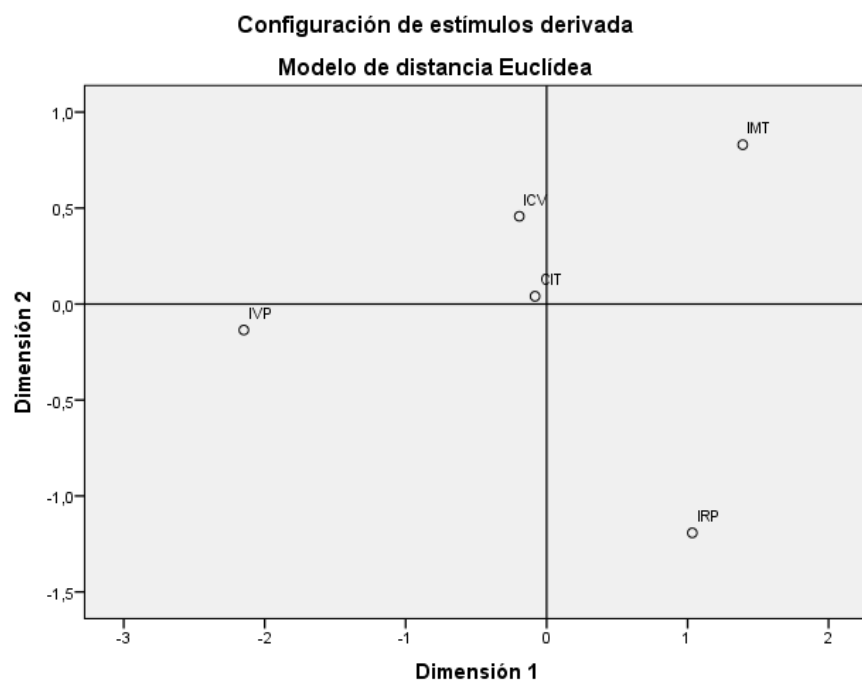
Escalamiento multidimensional pesado (WMDS)

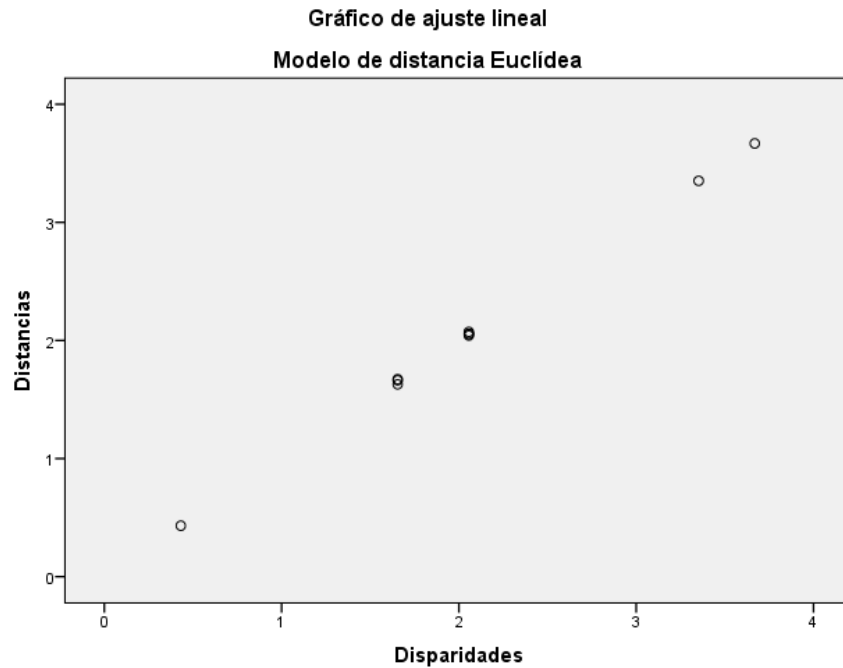
A continuación se muestran los resultados del escalamiento multidimensional pesado:

Índices del WMDS para las dos adaptaciones juntas: Stress = ,03752
RSQ = ,98498

Índices del WMDS para la adaptación mexicana: Stress = ,00555 RSQ = ,99979

Índices del WMDS para la adaptación estadounidense: Stress = ,00245
RSQ = ,96978





		1(Mx)	2(EEUU)
1	ICV	1,0346	,6479
2	IRP	-1,1883	,9302
3	IMT	-1,2281	-,2490
4	IVP	-,3553	-1,2282
5	CIT	1,7371	-,1008

Estas son las coordenadas de las dimensiones que tiene cada adaptación en relación con cada uno de los índices de la escala WISC-IV.

Se observa que la que mejor comportamiento tiene para nuestra población es la estadounidense.

Discusión

Para comenzar, la cantidad de sujetos que fueron tenidos en cuenta para el presente estudio fueron 76, los cuales fueron 26 hombres y 50 mujeres (gráfica 1) con edades entre los 6 años y dos meses y 16 años, 11 meses (gráfica 2). Se hizo una comparación entre las medidas de cada uno de los sujetos para cada índice de la inteligencia y con cada una de las adaptaciones, en estas comparaciones se observa que existe correlación alta (mayor a 0,9) entre las puntuaciones de los índices, así como también que existe diferencias significativas para cada uno de los grupos dado que $P < 0,05$, por lo tanto el nivel de significancia cae dentro de la zona de rechazo de la hipótesis nula que indica que no existen diferencias, por lo tanto se

acepta la hipótesis alterna y se concluye que sí existen diferencias entre cada uno de los índices para cada adaptación, donde la diferencia es positiva ubicándose al lado derecho; además de esto es importante ver que a pesar de que sí existen diferencias no es posible saber si estas son demasiado grandes o no, así como tampoco saber cuál de las dos sería más apta; entonces a partir de esto se realiza el WMDS permitiendo determinar el ajuste factorial de cada adaptación, y saber de acuerdo a la dimensión de las distancias cuál es más apta para la población del SAP.

De acuerdo a los índices de *stress* y *RSQ*, los cuales fueron de 0,3752 y 0,98498 se puede observar que sí tiene un ajuste apropiado y el modelo es bueno (*stress* cercano a 0), y el modelo explica la proporción de la variabilidad de los datos (*RSQ* cercano a 1).

Además con las coordenadas de las dimensiones, como los valores de la dimensión 2, que corresponde a la adaptación estadounidense, se puede determinar que es esta la que mejor se ajusta, mientras que la mexicana, a pesar de ser buena no se ajusta tan bien como la otra dado que las distancias son mayores.

Conclusión

Como conclusión general se puede decir que la adaptación estadounidense está mejor ajustada para la población colombiana que la adaptación mexicana, no obstante esta última también es posible utilizarla y no traería repercusiones, pues su ajuste también es muy bueno, pero no tanto como la estadounidense.

A pesar de los resultados obtenidos no es posible llegar a conclusiones que se puedan generalizar debido a que el tamaño del número total de parámetros que se están estimando (el número de coordenadas de los estímulos) es relativo al número de valores de datos de su matriz de datos, es decir, existen muy pocos datos para analizar, pues sólo los 76, y para este tipo de análisis es necesario tener más datos que permitan obtener resultados más confiables, y por eso no es posible estimar de forma precisa los valores de los parámetros.

Además de lo anterior, otro inconveniente estuvo en que la muestra de la población eran consultantes del SAP, por lo tanto las puntuaciones de los índices están relacionadas con algún diagnóstico particular, por lo tanto la distribución de la puntuación no es como se distribuye en la estandarización de la prueba, pues en este trabajo la media de cada índice estuvo alrededor de una puntuación de 85, indicando que los resultados son bajos y se ubican una desviación estándar por debajo de la media de la prueba; por lo tanto esto puede influenciar en los resultados por lo que tales puntuaciones son muy homogéneas y hay muy pocos valores que estén por encima de la media, y que permitan hacer un análisis más exacto.

Por último, aún si se pudiera hacer un análisis más exhaustivo, con mayores datos, una población más grande y aplicaciones independientes, no sería pertinente concluir cuál sería la más apta para usar en Colombia cuando lo mejor debería ser la realización de la adaptación y validación del WISC-IV para la población colombiana que permitiera usar este instrumento sin inconvenientes y teniendo un grupo de referencia más confiable.

Referencias

- Gómez-Benito, J. y Hidalgo-Montesinos, M. D. (2003) Desarrollos Recientes en Psicometría. Avances en Medición, 1 (1), pp. 17-36
- Guerrero, F.M. y Ramírez, J.M. (s.f.) El análisis de escalamiento multidimensional: una alternativa y un complemento a otras técnicas multivariantes. Departamento de Economía y Empresa Universidad Pablo de Olavide. España.
- Hambleton, R. K., Merenda, P. y Spielberg, C. (2005) Adapting Educational and Psychological Tests for Cross-Cultural Assessment. Taylor & Francis Group
- Hambleton, R. K. (1996) capítulo 6: Adaptación de test para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas en.: Muñiz, J. Psicometría. Editorial Universitas S.A.: Madrid, España.
- Muñiz, J., Elosua, P. y Hambleton, R.K. (2013) Directrices para la traducción y adaptación de los tests: segunda edición. Psicothema, Vol. 25, No. 2, 151-157
- Solano-Flores, G., Contreras-Niño, L. A. y Backhoff-Escudero, E. (2006). Traducción y adaptación de pruebas: Lecciones aprendidas y recomendaciones para países participantes en TIMSS, PISA y otras comparaciones internacionales. Revista Electrónica de Investigación Educativa, 8 (2).
- Balluerka, N., GorostiagaA., Alonso-Arbiol, I. y Haranburu, M. (2007) La adaptación de instrumentos de medida de unas culturas a otras: una perspectiva práctica. Psicothema, Vol. 19, nº 1, pp. 124-133. Universidad del País Vasco
- Hambleton, R. K. y Patsula, L. (1998) Adapting Tests For Use Inmultiple Languages And Cultures. Social Indicators Research, Vol. 45: 153–171, Kluwer Academic Publishers. Impreso en Holanda.
- Elosúa, P. y López, A. (1999) Funcionamiento diferencial de los ítems y sesgo en la adaptación de dos pruebas verbales Psicológica vol. 20, 23-40 Universidad del País Vasco
- Gaite. L., Ramírez, N., Herrera, S. Y Vázquez-Barquero J. L. (1997) Traducción y adaptación transcultural de instrumentos de evaluación en psiquiatría: aspectos metodológicos Archivos de Neurobiología; 60(2):91-111

Marco, I., González-Romá, V. y Gómez, J. (2000). Teoría de respuesta al ítem y análisis factorial confirmatorio: dos métodos para analizar la equivalencia psicométrica en la traducción de cuestionarios Vol. 12, Supl. nº 2, pp. 540-544 Psicothem

Anexos

Anexo 1: Directrices prácticas para la adaptación de los tests.

1. Directrices previas

DP1. Antes de comenzar con la adaptación hay que obtenerlos permisos pertinentes de quien ostente los derechos de propiedad intelectual del test.

DP2. Cumplir con las leyes y prácticas profesionales relativas al uso de tests que estén vigentes en el país o países implicados.

DP3. Seleccionar el diseño de adaptación de tests más adecuado.

DP4. Evaluar la relevancia del constructo o constructos medidos por el test en las poblaciones de interés.

DP5. Evaluar la influencia de cualquier diferencia cultural o lingüística en las poblaciones de interés que sea relevante para el test a adaptar.

2. Directrices de desarrollo

DD1. Asegurarse, mediante la selección de expertos cualificados, de que el proceso de adaptación tiene en cuenta las diferencias lingüísticas, psicológicas y culturales entre las poblaciones de interés.

DD2. Utilizar diseños y procedimientos racionales apropiados para asegurar la adecuación de la adaptación del test a la población a la que va dirigido.

DD3. Ofrecer información y evidencias que garanticen que las instrucciones del test y el contenido de los ítems tienen un significado similar en todas las poblaciones a las que va dirigido el test.

DD4. Ofrecer información y evidencias que garanticen que el formato de los ítems, las escalas de respuesta, las reglas de corrección, las convenciones utilizadas, las formas de aplicación y demás aspectos son adecuados para todas las poblaciones de interés.

DD5. Recoger datos mediante estudios piloto sobre el test adaptado, y efectuar análisis de ítems y estudios de fiabilidad y validación que sirvan de base para llevar a cabo las revisiones necesarias y adoptar decisiones sobre la validez del test adaptado.

3. Directrices de confirmación

DC1. Definir las características de la muestra que sean pertinentes para el uso del test, y seleccionar un tamaño de muestra suficiente que sea adecuado para las exigencias de los análisis empíricos.

DC2. Ofrecer información empírica pertinente sobre la equivalencia del constructo, equivalencia del método y equivalencia entre los ítems en todas las poblaciones implicadas.

DC3. Recoger información y evidencias sobre la fiabilidad y la validez de la versión adaptada del test en las poblaciones implicadas.

DC4. Establecer el nivel de comparabilidad entre las puntuaciones de distintas poblaciones por medio de análisis de datos o diseños de equiparación adecuados.

4. Directrices sobre la aplicación

DA1. Preparar los materiales y las instrucciones para la aplicación de modo que minimicen cualquier diferencia cultural y lingüística que pueda ser debida a los procedimientos de aplicación y a los formatos de respuesta, y que puedan afectar a la validez de las inferencias derivadas de las puntuaciones.

DA2. Especificar las condiciones de aplicación del test que deben seguirse en todas las poblaciones a las que va dirigido.

5. Directrices sobre puntuación e interpretación

DPI1. Interpretar las diferencias de las puntuaciones entre los grupos teniendo en cuenta la información demográfica pertinente.

DPI2. Comparar las puntuaciones entre poblaciones únicamente en el nivel de invarianza establecida para la escala de puntuación utilizada en las comparaciones.

6. Directrices sobre la documentación

DC1. Proporcionar documentación técnica que recoja cualquier cambio en el test adaptado, incluyendo la información y las evidencias sobre la equivalencia entre las versiones adaptadas.

DC2. Proporcionar documentación a los usuarios con el fin de garantizar un uso correcto del test adaptado en la población a la que va dirigido.