

Estructura y análisis diferencial del ítem de acorde a sexo y edad del Test de Inteligencia General Matrices en estudiantes universitarios

Structure and differential analysis of the item according to sex and age of the General Intelligence Test Matrices in university students

Cristian Ramos-Vera ¹
Universidad César Vallejo
Joe Jeremías Sáenz Torres ²
Universidad Nacional Mayor de San Marcos

Recibido: 16 - 04 - 21

Aceptado: 25 - 04 - 21

Publicado: 18 - 06 - 21

Resumen

La presente investigación tuvo por objetivo analizar las propiedades psicométricas del Test de Inteligencia General Matrices en estudiantes de cuatro universidades de la ciudad de Lima. Para ello, se trabajó en una muestra peruana no probabilística de 517 (347 varones y 170 mujeres) de 18 y 31 años. Se utilizó el método de ecuaciones estructurales (SEM) para la medición del instrumento, lo cual permitió confirmar la estructura de unidimensionalidad, de igual manera se realizó la invarianza del análisis diferencial de los ítems mediante modelo SEM y regresión logística binaria que señalan la equivalencia de medida. Asimismo, se propone trabajar una estructura reducida del instrumento, en lo que respecta a la medida del factor g, como propuesta de evaluación de la inteligencia, esto con el fin de adaptarse a la muestra en estudiantes jóvenes y adultos evaluados.

Palabras clave: Inteligencia general; factor g; estructura factorial; análisis diferencial del ítem.

Abstract

The aim of this research was to analyze the psychometric properties of the General Intelligence Matrices Test in students from four universities in the city of Lima. To this end, we worked on a nonprobabilistic Peruvian sample of 517 (347 males and 170 females) aged 18 and 31. The method of structural equations (SEM) was used to measure the instrument, which allowed the confirmation of the one-dimensional structure. Similarly, the invariance

1 Universidad César Vallejo. Lima, Perú.

Autor para correspondencia: cristony_777@hotmail.com ORCID: <https://orcid.org/0000-0002-3417-5701>

2 Docente, Universidad Nacional Mayor de San Marcos. Lima, Perú.

E-mail: jsaenz248@gmail.com ORCID: <https://orcid.org/0000-0002-1385-0691>

of the differential analysis of the items was carried out by means of a SEM model and binary logistic regression that indicates the equivalence of measurement. Likewise, it is proposed to work on a reduced structure of the instrument, regarding the measurement of the g factor, as a proposal for the evaluation of intelligence, this with the aim of adapting to the sample in young and adult students evaluated.

Keywords: General intelligence; g-factor, factor structure; differential analysis of the item.

La medida de la inteligencia, a lo largo de los años, ha sido un constructo clave en la psicología y en la evaluación psicológica, por lo que hacer una revisión de sus postulados es fundamental en los tiempos actuales (Sternberg, 2019), para eso, es necesario examinar de manera exhaustiva las pruebas de inteligencia utilizadas, tomando en cuenta la inmensa responsabilidad que tienen los profesionales que trabajan y administran dichos instrumentos (Sternberg, 2019).

La propuesta teórica del Test de Inteligencia General Matrices está analizada bajo el enfoque de los aportes de Charles Spearman, propuesta que ha logrado determinar la evaluación a través de un factor único, el factor G que explica el modelo factorial de la inteligencia a través de los elementos del Matrices (Sánchez et al., 2015; Walrath, Willis, Dumont & Kaufman, 2019). La presencia y el uso habitual de los ítems que presenta la prueba de Matrices, ya es algo constante en las pruebas de aptitudes mentales y de la capacidad cognitiva general (g), siendo una fuente importante de influencia en los test el Test Matrices Progresivas de Raven, considerada como una de las pruebas cognitivas de tipo no verbal de mayor uso en la evaluación clínica–educativa (Wongupparaj, Kumari & Morris, 2015).

La prueba de Matrices progresivas de Raven es uno de los instrumentos de mayor uso cuando se trabaja el concepto de inteligencia fluida en adultos (Sanchez-Sanchez & Santamaria, 2015). Uno de los principales atractivos que muestra el test es que los ítems estímulos que se presentan son visuales, con figuras y colores que permiten que las personas tomen interés en los elementos (Sanchez-Sanchez & Santamaria, 2015); al igual que el Matrices, sus elementos tienen una alta sensibilidad para evaluar la conducta inteligente (Sanchez-Sanchez, Santamaria & Abad, 2015).

Uno de los puntos clave del Matrices, es el hecho de trabajar bajo el enfoque de la evaluación de la inteligencia fluida, esta es la capacidad que tienen las personas para aprender desde los primeros años de escolarización hasta la etapa de la edad adulta, donde esta habilidad se consolida junto a las experiencias y nuevos conocimientos que permiten generar la inteligencia cristalizada (Horn, 1985; Walrath et al., 2019).

El test Matrices fue elaborado tomando en cuenta los aportes de la teoría de Spearman, quien consideró que las diversas puntuaciones que se obtienen, en

cualquier medida que se establezca de inteligencia, se podían dividir en dos componentes: a los que llamó factores, uno general o “g” que hace referencia al conocimiento en común que tienen las personas (Horn, 1985); y otro específico o “s” que hace referencia a las tareas específicas que cada persona posee en su quehacer diario (Horn, 1985; Walrath et al., 2019). También, consideró que cualquier medida de inteligencia podía presentar una proporción entre el factor g / s determinada, siendo así, que el presente test Matrices ha sido desarrollado con el fin de proporcionar una medida más eficiente de la capacidad cognitiva general (g) utilizando ítems no verbales en este formato (Sánchez-Sánchez & Santamaria, 2015).

Estas medidas presentan mayor saturación del factor g (Gignac, 2015; Myszkowski & Storme, 2018), por lo cual se desarrollaron versiones breves que faciliten su aplicación en menos tiempo y reporten con mayor rigor los elementos más representativos de la unidimensionalidad del factor general (Myszkowski & Storme, 2018), a partir de diferentes modelos estadísticos instrumentales robustos (Flores-Mendoza et al., 2018; García-Garzón, Abad & Garrido, 2019; Myszkowski, 2020; Partchev, 2020) brindando mayor evidencia a la medición del factor g.

Por otro lado, es importante mencionar el efecto Flynn, referido al aumento de la puntuación del coeficiente intelectual (IQ) en las medidas del formato matrices a través de la evaluación en la población adulta a nivel mundial durante más de cincuenta años (Wongupparaj et al., 2015). Esto afianza la necesidad de utilizar test más breves a causa del límite de tiempo y cansancio mental, lo cual no permitían estimar un correcto diagnóstico del IQ siendo reflejado en las diferencias de puntuaciones de décadas atrás con estudios más recientes (Flynn, 2012; 2013).

En los últimos años se han evaluado las diferencias según sexo a través de las Matrices Raven en universitarios con puntuaciones mayores a favor de los varones (Abdel-Khalek & Flynn, 2015; Bakhiet, Al-Qudah, Essa, Abdelrasheed, Cheng & Lynn, 2016; Essa, Abdelrasheed, Bakhiet, Cheng, Dwieb & Lynn, 2016; Lynn, Cheng, Choi, Cho & Lee, 2018). Asimismo, otras investigaciones refieren una mejor puntuación del nivel de inteligencia en adultos de mayor edad que en los más jóvenes (Pérez, 2018; Rossi-Casé, Doná, Biganzoli & Llanos-Borja, 2018). Además, diversos estudios latinoamericanos de las medidas Raven a pesar de reportar diferencias generales en las puntuaciones según género (Flores-Mendoza et al., 2018), no presentan otros análisis, que permitan determinar si es debido a un sesgo diferencial en varones o mujeres u otro grupo sociodemográfico en particular. Por lo cual es importante fortalecer la medición mediante análisis diferencial de los ítems (DIF, por sus siglas en inglés) del test Matrices, y verificar si las posibles diferencias puntuadas se deben a alguna característica específica de la muestra como el género (Mackintosh & Bennett, 2005; Waschl & Burns, 2020) o pertenecer a un grupo etario de la edad adulta (Hertzog, 2019).

A partir de lo anterior, la presente investigación se plantea el siguiente objetivo general: evaluar mediante el análisis factorial confirmatorio la unidimensionalidad del instrumento, y a través del análisis diferencial de los ítems de acorde al sexo y edad del test no verbal Matrices de Sanchez-Sanchez et al. (2015) en una muestra de estudiantes universitarios. Siendo de importancia, que hasta la fecha solo se cuentan con dos estudios comparativos en muestras de niños y adolescentes peruanos que utilizan el test Matrices Raven (Millones, Flores-Mendoza & Rivalles, 2015; Lynn, Cheng & Meisenberg, 2019), y no hay investigación alguna en universitarios peruanos. Por lo cual se procuró determinar las propiedades psicométricas, cumpliendo con los estándares para su uso concreto como los proporcionados por The Standards for Educational and Psychological Testing (American Educational Research Association, la American Psychological Association (APA) y el National Council on Measurement in Education (NCME, 2014).

MÉTODO

Diseño

Es una investigación de tipo psicométrico, ya que se estudia las características psicométricas de la prueba mediante el uso de un paradigma cuantitativo (Alarcón, 2008). Asimismo, se considera un estudio instrumental, porque se ajusta a los lineamientos de adaptación y propiedades psicométricas de un test (León & Montero, 2002).

Muestra

La muestra estuvo conformada por 517 estudiantes universitarios de cuatro instituciones superiores de Lima Metropolitana, 347 fueron mujeres y 170 fueron varones, que equivalen al 67.1% y al 32.9% respectivamente. Las edades de los participantes fueron entre 18 a 31 años (ME=21.70; DE=3.68), los cuales cursaban estudios de primer año hasta el quinto año universitario.

Tabla 1
Distribución de la muestra según sexo.

	f	%
Masculino	170	32.9
Femenino	347	67.1
Total	517	100

Nota. f: frecuencia

Instrumento

Se utilizó el Matrices, Test de Inteligencia General (Sánchez-Sánchez et al., 2015) que consta de 6 niveles con 36 ítems cada uno, que permiten adaptar la evaluación a diferentes grupos etarios. Evalúa la inteligencia general mediante un razonamiento

abstracto basado en matrices y, como resultado se obtiene un puntaje de aptitud (PA) y un índice general (IG). Presenta validez factorial y convergente. Asimismo, evidencia consistencia interna general de alfa de .86 y alfa ordinal de .93. Para la presente investigación se utilizó el nivel F, que corresponde a la edad de adultos, estudiantes universitarios. Se evaluó modelos de medición utilizando Análisis Funcional Confirmatorio (AFC) para probar diversos modelos. El primer modelo propuesto de 32 ítems después de eliminar cuatro ítems no significativos. El segundo modelo se basó en un proceso de parcelación, este consiste en un procedimiento estadístico que combina (suma o media de indicadores) para utilizarlo como variables observadas del factor latente (Little, 2013; Little, Rhemtulla, Gibson & Schoemann, 2013).

Procedimiento

Se solicitó permiso a los autores originales para la elaboración del estudio y se firmó un acuerdo de colaboración con TEA Ediciones, posteriormente se solicitó permiso a las universidades e instituciones educativas. Asimismo, respecto a la adaptación cultural, el instrumento no lo requiere porque es una prueba de razonamiento inductivo basado en estímulos no verbales (matrices) que engloba el grupo etario de 6 a 74 años. En esta misma línea, se coordinó con los encargados los horarios y aulas, asistiendo las fechas establecidas para la aplicación del instrumento en la ciudad de Lima Metropolitana.

La duración de aplicación de la versión original es de aproximadamente 45 minutos, sin embargo, para fines de la adaptación peruana, se realizó la aplicación durante 40 minutos, indicando las instrucciones generales en las universidades.

Para el procesamiento de la información, se tabuló mediante Microsoft Excel, donde se consignaron los datos de edad, sexo, las respuestas brindadas la forma de aplicación, que en este caso fue la forma F.

Análisis Estadístico

Se evaluó las modelos de medición utilizando el Análisis Funcional Confirmatorio (AFC) para probar diversos modelos. El primer modelo tenía 32 ítems después de eliminar cuatro ítems no significativos. El segundo modelo se basó en un proceso de parcelación, que combina (suma o media) indicadores para utilizarlo como variables observadas del factor latente (Little, 2013; Little, Rhemtulla, Gibson & Schoemann, 2013). Por lo tanto, se procedió a dividir cada nivel de la prueba de Matrices (A, B, C) en dos grupos, manteniendo el mayor número de ítems en cada parcela, se obtuvo un conjunto preciso de estimaciones de parámetros de acorde a Little, Cunningham, Shahar y Widaman (2002) y se normalizó la tendencia de distribución de puntuaciones (Asparouhov, 2005; Bandalos & Finney, 2001). Se tuvo en cuenta el índice de dificultad que se obtuvo mediante la división del número de evaluados que respondió correctamente el ítem entre el número total de la muestra (Finch & French, 2019). Debido a que los ítems finales de cada nivel

eran más complejos y las parcelas se ajustan mejor al mayor número de ítem, se procedió a parcelar los primeros ítems de cada nivel. Se obtuvo así 6 parcelas (2 por nivel) en donde los elementos dicotómicos se resumen de la siguiente manera: $A1+A2+A3+A4+A5+A6=$ Parcela 1, $A7+A8+A9+A10+A11+A12=$ Parcela 2, $B1+B2+B3+B4+B5+B6=$ Parcela 3 y así sucesivamente. Mediante este procedimiento al combinar el mismo número de ítems se mantienen la varianza compartida. Una parcela determinada tendrá una mayor proporción de varianza verdadera (Little, 2013; Little et al., 2013). En consecuencia, cuanto mayor sea el número de indicadores sumados mejor proporción de la varianza de la puntuación real. Se evaluó los supuestos de distribución de los datos de asimetría y curtosis ± 1.5 (Bandalos & Finney, 2010) además se realizó el coeficiente de Mardia considerando la normalidad general a partir de valores ≤ 70 (Mardia, 1974). Finalmente se realizó un modelo breve utilizando los 24 primeros ítems que representan las cuatro parcelas con cargas factoriales mayor a .50, posteriormente se retuvo 18 reactivos que presentaron carga factorial mayor a .35 (Brown, 2015) para evaluar la medida breve del factor g. Estos 18 elementos (A2, A3, A4, A6, A7, A8, A9, A10, B1, B2, B3, B4, B6, B7, B8, B9, B11, B12) se cargan en un único factor latente mediante el AFC como una versión reducida de la prueba y propuesta para la evaluación de inteligencia en adultos.

Con relación a la bondad del ajuste, se utilizó el ajuste comparativo (CFI), el índice de Tucker-Lewis (TLI), el error cuadrático medio de aproximación (SRMR) y la media de la raíz aproximación de error cuadrado (RMSEA). Los criterios de análisis utilizados para determinar la bondad de ajuste se describen a continuación: RMSEA y SRMR menores de .08, mientras que el CFI y TLI mayores de .90 (Hu y Bentler, 1999). CFI y TLI son penalizados bajo modelos complejos (es decir, modelos multidimensionales con muchos ítems por factor) y tienden a empeorar a medida que aumenta el número de variables en el modelo (Kenny & McCoach, 2003). Por lo tanto, como se ha indicado anteriormente considerando el número de ítems, los valores cercanos a 0,9 también pueden considerarse de buen ajuste, se utilizó el método de estimación *Diagonally Weighted Least Squares* (DWLS), dado que las variables utilizadas son dicotómicas (0, 1), que corresponden a las variables categoriales (Li, 2016), este análisis es recomendable teniendo en cuenta la naturaleza de los elementos dicotómicos (Gana & Broc, 2019; Little, 2013).

Se consideró la prueba de invarianza de medición a través del método de múltiples indicadores de causa múltiples “MIMIC” (Barendse, Oort, Werner, Ligtvoet & Schermelleh-Engel, 2012; Finch, 2005; Shih & Yang, 2009) en medidas dicotómicas (por ejemplo, Baños & Ramos-Vera, 2020). Después de realizar ese modelo se propuso el modelo MIMIC con los respectivos ítems con DIF en las covariables de sexo y la edad.

Los efectos directos entre sexo y edad fueron incluidos hacia los 18 indicadores respectivos del modelo 4 uno por uno, y solo fueron retenidas aquellas rutas

significativas que refieren posible evidencia DIF, según el proceso iterativo "MIMIC model with scale purification" (M-SP, Wang & Shih, 2010).

Posteriormente se realizó el análisis similar de detección del DIF uniforme y no uniforme teniendo en cuenta el procedimiento de regresión logística binaria (Narayanan & Swaminathan, 1996; Swaminathan & Rogers, 1990). Este método permite comparar estadísticamente un modelo en el que se predice la posibilidad de acertar un ítem de la puntuación total con un modelo en el que adicionalmente se añaden como predictores el grupo de pertenencia y la interacción entre la puntuación total y el grupo, siendo otro proceso de DIF como el modelo MIMIC en ítems dicotómicos (Gelin & Zumbo, 2007; Lee, Bulut & Suh, 2016). El análisis de regresión logística se realizó mediante el módulo PsychoPDA Bynary LogR (Friesen, Kroc & Zumbo, 2019). Se tuvo en consideración el tamaño de efecto más frecuente en este tipo de análisis, la varianza contabilizada ΔR^2 (Sireci & Ríos, 2013), basado en la diferencia entre los coeficientes R^2 de Nagelkerke (Nagelkerke, 1991) de ambos modelos utilizando el estimador estadístico de la prueba de razón de verosimilitud (Fidalgo, Alavi & Amirian, 2014; Lee et al., 2016). Según la propuesta de Thomas y Zumbo (1996), y Zumbo (1999) donde el tamaño de efecto DIF es pequeño ($0 < \Delta R^2 < .13$), moderado ($.13 < \Delta R^2 < .26$) o grande ($\Delta R^2 > .26$).

RESULTADOS

Análisis preliminar de los ítems

En la Tabla 2 se presentan los estadísticos descriptivos y de distribución para cada ítem del Matrices, Test de inteligencia general. En este sentido, se evidencia que los valores de asimetría y curtosis se encuentran cercanos o superior a +/- 1.5 (Bandalos & Finney, 2010).

Tabla 2
Análisis descriptivo de los ítems

<i>Ítem</i>	<i>M</i>	<i>DE</i>	<i>g₁</i>	<i>g₂</i>
1	.91	.28	-2.92	6.53
2	.85	.36	-1.95	1.80
3	.86	.35	-2.04	2.17
4	.85	.35	-2.00	2.03
5	.86	.35	-2.11	2.44
6	.81	.40	-1.54	0.38
7	.77	.42	-1.31	-.28
8	.80	.40	-1.49	0.23
9	.68	.47	-.76	-1.42
10	.65	.48	-.65	-1.58
11	.49	.50	.05	-2.00
12	.65	.48	-.62	-1.61
13	.70	.46	-.87	-1.24

Tabla 2. *Continuación*

<i>Ítem</i>	<i>M</i>	<i>DE</i>	<i>g₁</i>	<i>g₂</i>
14	.72	.45	-.98	-1.03
15	.63	.48	-.53	-1.72
16	.69	.46	-.83	-1.30
17	.62	.49	-.50	-1.75
18	.70	.46	-.90	-1.20
19	.39	.49	0.44	-1.81
20	.41	.49	0.36	-1.87
21	.51	.50	-.020	-2.00
22	.40	.49	0.41	-1.84
23	.31	.46	0.82	-1.33
24	.40	.49	0.42	-1.82
25	.36	.48	0.58	-1.66
26	.36	.48	0.60	-1.64
27	.31	.46	0.84	-1.30
28	.38	.49	0.49	-1.77
29	.20	.40	1.50	0.27
30	.21	.41	1.43	0.04
31	.21	.41	1.34	-.040
32	.19	.39	1.60	0.57
33	.27	.46	1.62	5.78
34	.18	.38	1.67	0.78
35	.22	.42	1.33	-.21
36	.30	.46	0.89	-1.21

Nota: *M* = Media; *DE* = Desviación estándar; *g₁* = Asimetría; *g₂* = Curtosis.

En la tabla 3 se muestra un mayor índice de dificultad en los ítems de A1 hasta B12 que refieren la presencia del factor general.

Tabla 3

Índice de Dificultad de los 36 ítems de la Prueba Matrices

Item	Dificultad	Item	Dificultad	Item	Dificultad
A1	0.768	B1	0.478	C1	0.217
A2	0.602	B2	0.468	C2	0.242
A3	0.638	B3	0.625	C3	0.259
A4	0.679	B4	0.588	C4	0.199
A5	0.723	B5	0.462	C5	0.215
A6	0.609	B6	0.532	C6	0.155
A7	0.569	B7	0.284	C7	0.168
A8	0.638	B8	0.337	C8	0.193
A9	0.482	B9	0.340	C9	0.149
A10	0.441	B10	0.284	C10	0.166
A11	0.511	B11	0.240	C11	0.155
A12	0.389	B12	0.352	C12	0.099

Nota: A: Nivel a, B: Nivel b, Nivel c.

Confiabilidad

El estudio de confiabilidad se realizó a partir del método de consistencia interna, por medio del Coeficiente Omega ω . En la Tabla 4 se observa que, para las formas del Matrices, Test de Inteligencia General, compuesto por 36 ítems, se obtuvo valores de .49 a .71, y una consistencia general del instrumento de .80.

Tabla 4
Fiabilidad del Matrices, Test de inteligencia general

Nivel	Coefficiente Omega ω
A	.71
B	.70
C	.49
General	.80

Nota. A: Nivel a, B: Nivel b, C: Nivel c.

Análisis factorial Confirmatorio (AFC)

Se realizó 4 modelos estructurales en el presente (tabla 5). El modelo M1 representa la estructura unidimensional de 33 ítems respectivos (se eliminó 3 ítems, los ítems 32, 33 y 35 que no eran significativos en el modelo unidimensional), este modelo evidenció un ajuste de proporción con valores adecuados ($\chi^2/gl=1.37$, CFI=.88, TLI=.88, SRMR=.043, RMSEA=.027). En el modelo M2 de 6 parcelas respectivas de 6 ítems cada uno presenta adecuados índices de ajuste ($\chi^2/gl=3.60$, CFI=.96, TLI=.93, SRMR=.046, RMSA=.072). Este modelo de inteligencia general presenta una distribución de curtosis de Mardia de $z=1.775$ que indica una distribución adecuada de normalidad multivariada del modelo propuesto. También se evaluó un modelo unidimensional (modelo 3) breve de 18 ítems con cargas factoriales mayores de .35 (Brown, 2015) con adecuados valores de ajuste ($\chi^2/gl=1.40$, CFI=.95, TLI=.95, SRMR=.038, RMSA=.028). Finalmente, para los resultados de invariabilidad se tuvo en consideración el modelo anterior de 18 ítems (modelo 3). Los 18 indicadores del modelo breve fueron evaluados mediante el método MIMIC (modelo 4) presentan índices de ajuste propicios ($\chi^2/gl=1.28$, CFI=.96, TLI=.95, SRMR=.037, RMSA=.024), con respecto al covariable sexo, si presentó efecto directo en el factor latente de inteligencia al igual que en el indicador B12 ($\beta=-.11$, $p=.009$). Esto indica que cuando el factor g se mantiene constante hay un posible efecto de diferencia según sexo ($\beta=-.13$, $p=.011$) donde un efecto negativo favorece al primer grupo y una predicción positiva al segundo (1 para mujeres y 2 para varones), que indicaría haber una tendencia mayor de respuesta en las mujeres (Cheng, Chao & Lathrop, 2016). En relación a la covariable edad se dividió en dos grupos al determinar el valor de la mediana muestral de 20 (1 = 18-20 años, 2 = 21-31 años), esta covariable no

presentó efecto significativo en la variable latente mientras que si predijo significativamente en los ítem A8, B7, B8 y B9, con valores de: A8 ($\beta = -.10, p = .012$), B7 ($\beta = -.12, p = .006$), B8 ($\beta = -.10, p = .022$) y B9 ($\beta = -.16, p = .000$), lo cual señala un sesgo cuestionable que favorece a los evaluados menores de 21 años debido a que no cumplió con la significancia a nivel latente, aun así no se puede estar seguro de los efectos DIF en ambos grupos según edad en los ítems respectivos (Brown, 2015).

Tabla 5

Parámetros de los ítems: modelo unifactorial, de parcelación y modelo breve de 16 ítems, modelo MIMIC de 16 ítems (n=517)

M	X ² /gl	P	CFI	TLI	SRMR	RMSEA
M1	1.37	<.001	.884	.876	.043	.027
M2	3.60	<.001	.960	.934	.046	.072
M3	1.40	<.001	.953	.946	.038	.028
M4	1.28	.006	.959	.952	.037	.024

Nota: M1= Modelo unifactorial de 33 ítems, M2= modelo de parcelación unifactorial en 6 facetas de los 36 ítems, M3= Modelo del factor general de 18 ítems, M4= Modelo MIMIC del factor general de 18 ítems, x²/gl = índice de Chi cuadrado relativo, p = significancia, CFI = índice de bondad de ajuste comparativo, TLI = índice de Tucker Lewis, RMSEA = raíz cuadrada del error medio cuadrático, SRMR = raíz residual estandarizada cuadrática media.

ANÁLISIS DE REGRESION LOGISTICA PARA DIF

Posteriormente se realizó el análisis de regresión logística para determinar el tamaño de efecto DIF uniforme y no uniforme según sexo y edad para confirmar la magnitud de los efectos DIF según el modelo MIMIC-SEM de la medida breve. Se obtuvo valores DIF mínimos de nivel A que según la escala de Zumbo y Tomas (1996) se pueden interpretar como valores insignificantes que no determinan sesgo a nivel de ítem del instrumento general Matrices.

Según la tabla 6 se presentan valores DIF según edad de nivel A según la escala de Tomas y Zumbo (1996), el ítem A8 refiere el menor tamaño DIF ($\Delta R^2 = 0.013$), mientras que el elemento C10 reportó el mayor valor DIF ($\Delta R^2 = 0.038$), con respecto a la agrupación por sexo los niveles DIF obtenidos son B12 ($\Delta R^2 = 0.017$) y C11 ($\Delta R^2 = 0.035$). Por lo tanto, estos resultados también permiten interpretar que los efectos DIF son insignificantes, teniendo en cuenta al umbral B (.13-.26) como el tamaño de efecto mínimo de interés, por ende, no presentan sesgo diferencial en los ítems de la medida breve del Matrices (Finch & French, 2019; Friesen et al., 2019; Zumbo, 1999).

Tabla 6
Análisis diferencial del ítem (DIF) mediante Regresión Logística Binaria

Ítem	DIF Según edad				Ítem	DIF según sexo			
	ZT	P	X ²	Δ R ²		ZT/JG	P	X ²	Δ R ²
A8	NIVEL A	.037	6.62	.013	B6	NIVEL A	.005	10.63	.020
A12	NIVEL A	.006	10.25	.024	B12	NIVEL A	.023	7.56	.017
B3	NIVEL A	.040	6.43	.014	C11	NIVEL A	.005	10.58	.035
B7	NIVEL A	.007	9.89	.022	-	-	-	-	-
B8	NIVEL A	.034	6.75	.016	-	-	-	-	-
B9	NIVEL A	.001	15.19	.035	-	-	-	-	-
C2	NIVEL A	.035	6.70	.018	-	-	-	-	-
C3	NIVEL A	.013	8.76	.022	-	-	-	-	-
C4	NIVEL A	.015	8.39	.025	-	-	-	-	-
C6	NIVEL A	.006	10.15	.032	-	-	-	-	-
C10	NIVEL A	.002	12.64	.038	-	-	-	-	-

Nota_ ZT: escala de Zumbo Tomas, p: nivel de significancia, x2: prueba de chi cuadrado, Δ R²: Tamaño de efecto

DISCUSIÓN

La medición equitativa del razonamiento del nivel de inteligencia general de los universitarios señala uno de los mayores determinantes para el éxito profesional y académico (Danner et al., 2016; Wicherts, 2016). A partir de la evaluación de las propiedades psicométricas del Test de Inteligencia Matrices se determinó el modelo unidimensional propuesto por los autores originales y la invariancia de medición según sexo y edad. Los resultados apoyan los modelos mencionados. Además, mediante el análisis MIMIC se obtuvo la invariancia en su mayoría de ítems a excepción de seis ítems, luego al mediante el DIF basado en regresión logística binaria se determinó la insignificancia de los efectos obtenidos.

Los resultados del modelo 1 de 32 ítems mostraron un nivel próximo a un buen ajuste de acuerdo con el modelo unidimensional original propuesto por Sanchez-Sanchez et al. (2015), Se tuvo en consideración el posicionamiento de los ítems según los niveles para realizar el análisis de parcelación propuesto en el modelo 2 que demostró evidencias adecuadas teniendo en consideración los 36 ítems. Se encontró una mayor dificultad en los reactivos A12, B11 y C12.

También se puede observar en la tabla 2 que el aumento de dificultad característico de estas pruebas de inteligencia relacionados al RAVEN muestra una tendencia general de interrupción, este fenómeno es más notorio en los elementos A5 y B3. De acorde a las cargas factoriales del modelo estructural agrupado en parcelas se determinó mayores coeficientes estandarizados en los niveles A y B

(.66, .75, .69, y .59; ítems de A1-A6, A7-A12, B1-B6, B7-B12, respectivamente) en comparación al nivel C (.41 y .18). Esto indicaría que los reactivos del nivel A y B son menos complejos que los reactivos del nivel C en términos de contenido para esta muestra generando resultados fiables. Por lo tanto, se realizó el Modelo 3 para evaluar estos 24 reactivos de las cuatro parcelas respectivas, de los cuales se retiró 6 ítems (A1, A5, A11, A12, B5, B10) que no se adecuaron al modelo unidimensional propuesto. El análisis con los 18 elementos restantes presentó buen ajuste estructural.

Posteriormente se realizó el análisis de invariancia MIMIC según la edad, que refieren que cuatro ítems presentaban DIF (ítems A8, B7, B8 y B9) mediante un patrón uniforme, pues estos indicadores tienden a tener una mayor probabilidad de ser realizadas por los evaluados menores de 21 años, aun así, la variable categórica reporta un efecto no significativo sobre las puntuaciones medias de la inteligencia ($p=.078$). Mientras que el análisis MIMIC DIF de acorde al sexo fue significativo en el ítem B12, asimismo, señala un impacto significativo de la variable categórica en el factor general de inteligencia indicando diferencias en dicho reactivo a favor de las mujeres. Estos valores a través del método SEM presentaron coeficientes estandarizados de tamaño de efecto menor al mínimo requerido de $\pm .20$ (Finch & French, 2019) por lo cual no se puede confirmar la presencia de DIF en el modelo propuesto.

Finalmente, para corroborar los resultados anteriores, se realizó el modelo de regresión logística binaria, que evidenció un tamaño de efecto DIF mínimo (de nivel A; Zumbo, 1999) en 14 reactivos (A8, A12, B3, B6, B7, B8, B9, B12, C2, C3, C4, C6, C10 y C11), que pueden interpretarse como un sesgo insignificante, determinando la falta equivalencia de todos los reactivos del Test Matrices de acorde a los covariables de edad y sexo (Friesen et al., 2019; Uğurlu & Burcu, 2020). Un aspecto por considerar es que el método MIMIC presenta como limitación la falta de detección de DIF no uniforme a diferencia del análisis de regresión logística (Gómez-Benito, Hidalgo & Padilla, 2009; Jamali, Ayatollahi & Jafari, 2016).

Asimismo, según la regresión logística también se evidencia un nivel mínimo DIF uniforme y no uniforme. Lo cual permite interpretar que los resultados del estudio respaldan la invarianza de la investigación instrumental original en adultos (Sanchez-Sanchez et al., 2015). Este análisis fortalece la interpretación de los resultados con respecto al primer análisis DIF, pues se han demostrado posibles efectos elevados en modelos MIMIC SEM (Woods & Grimm, 2011).

Dentro de las limitaciones y fortalezas de este estudio. Se considera una limitación que los resultados refieren solo en el nivel F (muestra en adultos) dentro del rango de 18 a 31 años. Estos resultados no se pueden generalizar en muestras de los diferentes niveles de edad propuesto en el estudio original (Sanchez-Sanchez et al., 2015). El presente estudio es importante por una serie de razones. Es la primera

adaptación del instrumento MATRICES que carece de estudios que exploren sus propiedades psicométricas. Este estudio es pionero en la adaptación de este instrumento en el contexto latinoamericano basado en los test matrices “RAVEN” conocido como una de las pruebas para evaluar la inteligencia más utilizadas a nivel mundial (Wongupparaj et al., 2015). A pesar de que solo se cuenta con la muestra original en adultos, los resultados están de acorde a la investigación original del Matrices (por ejemplo, unidimensionalidad e invariabilidad en el factor g según sexo y edad). Estos resultados refuerzan la robustez de la validez del constructo de la habilidad no verbal. También proponemos un conjunto reducido de elementos que pueden utilizarse para evaluar el factor g en adultos con un índice de ajuste adecuado. Dichas características psicométricas deben ser evaluadas en diferentes muestras de validación para corroborar los resultados obtenidos. Se recomiendan evaluar otros métodos estadísticos que refuercen las evidencias de los hallazgos como el factor Bayes (Ramos-Vera, 2020) y el análisis de redes (García-Garzon, Abad & Garrido, 2019; Ramos-Vera, 2021).

CONCLUSIONES

El presente estudio psicométrico de Matrices fue realizado en una muestra de adultos. Los resultados avalan los hallazgos de la unidimensionalidad en adultos de acorde al estudio español inicial. También se demostró que los elementos del instrumento son invariables. Proponemos una estructura reducida instrumental del *factor g* que se adapta a la muestra en adultos. Las investigaciones futuras deben validar de forma cruzada los hallazgos para determinar niveles más sólidos de invariabilidad y validez del factor g principal de la versión breve. También se debe tener en cuenta realizar estudios psicométricos con respecto a los diferentes niveles de edad del presente instrumento.

Conflictos de interés

Los autores declaran que no existen conflictos de interés en la elaboración del manuscrito.

Financiamiento

Autofinanciado.

REFERENCIAS

Abdel-Khalek, A. M., & Lynn, R. (2016). Sex Differences in the Intelligence of Students at Alexandria University, Egypt. *Mankind Quarterly*, 57(1), 87-90. Recuperado de https://www.researchgate.net/publication/309104195_Sex_differences_in_the_intelligence_of_students_at_Alexandria_University_Egypt

- AERA, APA, NCME. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association, *American Psychological Association*, National Council on Measurement in Education.
- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 411-434. doi:10.1207/s15328007sem1203_4
- Bakhiet, S. F. A., Al-Qudah, M. F., Essa, Y. A. S., Abdelrasheed, N. S. G., Cheng, H., & Lynn, R. (2016). Sex Differences in the Intelligence of University Engineering Students in Sudan. *Mankind Quarterly*, *57*(1), 95-98. Recuperado de: <https://mankindquarterly.org/archive/issue/57-1/13>
- Bandalos, D., & Finney, S. (2001). Item parceling issues in structural equation modeling. En G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Hillsdale, NJ: Lawrence Erlbaum.
- Bandalos, D., y Finney, S. (2010). Factor Analysis: Exploratory and Confirmatory. En G. R. Hancock y R. O. Mueller (Eds.), *Reviewer's guide to quantitative methods*. Routledge: New York.
- Baños-Chaparro, J., Ramos-Vera, C. Validez e invarianza según sexo y edad de la Escala Paykel de Ideación Suicida en adolescentes peruanos. *Interacciones*. 2020. 6(1), e225. <https://doi.org/10.24016/2020.v6n1.225>
- Barendse, M. T., Oort, F. J., Werner, C. S., Ligtvoet, R., & Schermelleh-Engel, K. (2012). Measurement Bias Detection Through Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(4), 561-579. doi:10.1080/10705511.2012.713261
- Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507-526. Obtenido de: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5140785/?_escaped_fragment_&po=38.3333
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, NY: Guilford Publications
- Cheng, Y., Shao, C., & Lathrop, Q. N. (2015). The Mediated MIMIC Model for Understanding the Underlying Mechanism of DIF. *Educational and Psychological Measurement*, *76*(1), 43-63. doi:10.1177/0013164415576187
- Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D. L., ... Ziegler, M. (2016). Current challenges, new developments, and future directions in scale construction. *European Journal of Psychological Assessment*. *32*, 175-180. doi: 10.1027/1015-5759/a000375
- Essa, Y. A. S., Abdelrasheed, N. S. G., Bakhiet, S. F. A., Cheng, H., Dwieb, A. M. M., & Lynn, R. (2016). Sex differences in the intelligence of students at an Egyptian university. *Personality and Individual Differences*, *95*, 183-184. <https://doi.org/10.1016/j.paid.2016.02.036>
- Fidalgo, A. M., Alavi, S. M., & Amirian, S. M. R. (2014). Strategies for testing statistical and practical significance in detecting DIF with logistic regression models. *Language Testing*, *31*(4), 433-451. doi:10.1177/0265532214526748

- Finch, H. (2005). The MIMIC Model as a Method for Detecting DIF: Comparison With Mantel-Haenszel, SIBTEST, and the IRT Likelihood Ratio. *Applied Psychological Measurement, 29*(4), 278–295. doi:10.1177/0146621605275728
- Finch, H., y French, B. (2019). *Educational and Psychological Measurement*. New York, NY: Taylor & Francis.
- Flores-Mendoza, C., Ardila, R., Rosas, R., Lucio, M. E., Gallegos, M., & Reátegui-Cola-reta, N. (2018). *Intelligence, Latin America, and Human Capital. Intelligence Measurement and School Performance in Latin America*, 79–112. doi:10.1007/978-3-319-89975-6_6
- Flynn, J. R. (2012). *Are we getting smarter: Rising IQ in the twenty-first century*. Cambridge: Cambridge University Press.
- Flynn, J. R. (2013). The “Flynn Effect” and Flynn's paradox. *Intelligence, 41*(6), 851–857. <https://doi.org/10.1016/j.intell.2013.06.014>
- Friesen, L., Kroc, E., Zumbo, B. (2019). *Psychometrics & Post-Data Analysis: Binary Differential Item Functioning*. [Jamovi module]. Retrieved from <https://github.com/lucasjfriesen/jamoviPsychoPDA>.
- Gana, K., y Broc, G. (2019). *Structural Equation Modeling with Lavaan*. New York: Wiley.
- García-Garzon, E., Abad, F. J., & Garrido, L. E. (2019). Searching for g: A new evaluation of spm-ls dimensionality. *Journal of Intelligence, 7*(3), 14. <https://doi.org/10.3390/jintelligence7030014>
- Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence, 52*, 71–79. doi: 10.1016/j.intell.2015.07.006
- Gómez-Benito, J., Hidalgo, M. D., & Padilla, J.L. (2009). Efficacy of effect size measures in logistic regression: An application for detecting DIF. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 5*(1), 18–25. <https://doi.org/10.1027/1614-2241.5.1.18>
- Hertzog, C. (2019). Intelligence in Adulthood. En R.J. Sternberg (Ed). *The Cambridge Handbook of Intelligence*, (pp. 181–204). Cambridge: Cambridge University Press. doi:10.1017/9781108770422.010
- Horn, J. L. (1985). Remodeling old models of intelligence. En B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 267-300). Nueva York: Wiley.
- Hu, L. T. & Bentler, P. M. (2016). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi:10.1080/10705519909540118
- Jamali, J., Ayatollahi, S. M., & Jafari, P. (2016). A New Measurement Equivalence Technique Based on Latent Class Regression as Compared with Multiple Indicators Multiple Causes. *Acta informatica medica: AIM: journal of the Society for Medical Informatics of Bosnia & Herzegovina: casopis Drustva za medicinsku informatiku BiH, 24*(3), 168–171. doi:10.5455/aim.2016.24.168-171

- JAMOVI project (2019). jamovi. (Version 1.0) [Computer Software]. Recuperado de: <https://www.jamovi.org>.
- JASP Team (2019). JASP (Versión 0.11) [Computer software].
- Kenny, D., & McCoach, D. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10, 333-351. doi:10.1207/S15328007SEM1003_1
- Lee, S., Bulut, O., & Suh, Y. (2016). Multidimensional Extension of Multiple Indicators Multiple Causes Models to Detect DIF. *Educational and Psychological Measurement*, 77(4), 545–569. doi:10.1177/0013164416651116
- Li, C.H. (2016) Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research Methods*, 48(3), 936-949. doi: 10.3758/s13428-015-0619-7
- Little, T. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.
- Little, T., Cunningham, W., Shahar, G., & Widaman, K. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173. doi:10.1207/S15328007SEM0902_1
- Little, T., Rhemtulla, M., Gibson, K., & Schoemann, A. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18(3), 285–300. doi:10.1037/a0033266
- Lynn, R., Cheng, H., Choi, Y. Y., Cho, S. H., & Lee, K. H. (2018). Sex Differences in Intelligence of University Students in South Korea. *Mankind Quarterly*, 59(1), 108-114. Recuperado de: <https://mankindquarterly.org/archive/issue/59-1/9>
- Lynn, R., Cheng, H., & Meisenberg, G. (2019). Cognitive Abilities in Young Lives: An Overview of Results from Ethiopia, India, Peru and Vietnam. *Mankind Quarterly*, 59(3), 406-422. Obtenido de: <http://emilkirkegaard.dk/en/wp-content/uploads/Lynn-2019-young-lives.pdf>
- Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33(6), 663-674. <https://doi.org/10.1016/j.intell.2005.03.004>
- Mardia, K. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya*, B36, 115–128. Recuperado de: <https://www.jstor.org/stable/25051892?seq=1>
- Millones, D., Flores-Mendoza, C., & Rivalles, R. (2015). Intelligence in Peru: Students' results in Raven and its relationship to SES. *Intelligence*, 51, 71–78. doi: 10.1016/j.intell.2015.05.004
- Myszkowski, N. (2020). A Mokken Scale Analysis of the Last Series of the Standard Progressive Matrices (SPM-LS). *Journal of Intelligence*, 8(2), 22. doi:10.3390/jintelligence8020022

- Myszkowski, N., & Storme, M. (2018). ¿A snapshot of g? Binary and polytomous item-response theory investigations of the last series of the Standard Progressive Matrices (SPM-LS). *Intelligence*, *68*, 109-116. <https://doi.org/10.1016/j.intell.2018.03.010>
- Nagelkerke, N. (1991). A note a general definition on the coefficient of determination. *Biometrika*, *78*, 691-692. Recuperado de: <https://pdfs.semanticscholar.org/1970/6b6e9ba4050a20f2980bealde35d23882b51.pdf>
- Narayanon, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, *20*(3), 257-274. Recuperado de: <https://conservancy.umn.edu/bitstream/handle/11299/119459/v20n3p257.pdf?sequence=1>
- Partchev, I. (2020). Diagnosing a 12-Item Dataset of Raven Matrices: With Dexter. *Journal of Intelligence*, *8*(2), 21. <https://doi.org/10.3390/jintelligence8020021>
- Pérez, M.-A. (2018). Baremos del TMP Raven Escala General. X Congreso Internacional de Investigación y Práctica Profesional en Psicología XXV Jornadas de Investigación XIV Encuentro de Investigadores en Psicología del MERCOSUR. Facultad de Psicología, Universidad de Buenos Aires, Buenos Aires, Argentina.
- Ramos-Vera, C. A. (2020). La alternativa metodológica del factor Bayes en la investigación clínica de nutrición. *Revista Española de Nutrición Humana y Dietética*, *24*(4), 401-403. <http://dx.doi.org/10.14306/renhyd.24.4.1231>
- Ramos-Vera, C. (2021). Las redes de relación estadística en la investigación psiquiátrica: el caso del delirio en el contexto de COVID-19. *Revista Colombiana De Psiquiatría*. <https://doi.org/10.1016/j.rcp.2021.02.004>
- Rossi-Casé, L., Maris-Doná, S., Garzaniti, R., Biganzoli, B., & Llanos-Barja, C. (2018). La inteligencia a través de las generaciones: Millennials y centennials. *Acta de investigación psicológica*, *8*(2), 90-100. <http://dx.doi.org/10.22201/fpsi.20074719e.2018.2.08>
- Sánchez-Sánchez, F., y Santamaría, P. (2015). La evaluación inteligente de la inteligencia. *Padres y Maestros/Journal of Parents and Teachers*, *364*, 55-63. <https://doi.org/10.14422/pym.i364.y2015.009>
- Sánchez-Sánchez, F., Santamaría, P. y Abad, F. J. (2015). *Matrices. Test de Inteligencia General*. Madrid: TEA Ediciones.
- Shih, C.L., & Wang, W.C. (2009). Differential item functioning detection using the multiple indicators, multiple causes method with a pure short anchor. *Applied Psychological Measurement*, *33*(3), 184-199. doi: 10.1177/0146621608321758
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation*, *19*(2-3), 170-187. doi:10.1080/13803611.2013.767621
- Sternberg, R.J. (2019). *The Cambridge Handbook of Intelligence*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108770422.004>
- Swaminathan, H., & Rogers, H. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, *27*(4), 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

- Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral Statistics, 21*(2), 110–130. doi:10.3102/10769986021002110
- Uğurlu, S., & Burcu, A. T. A. R. (2020). Performances of MIMIC and Logistic Regression Procedures in Detecting DIF. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 11*(1), 1-12. Recuperado de: <https://dergipark.org.tr/en/download/article-file/920579>
- Yang, C., Nay, S. y Hoyle, R. (2010). Three Approaches to Using Lengthy Ordinal Scales in Structural Equation Models. *Applied Psychological Measurement, 34*(2), 122–142. doi:10.1177/0146621609338592
- Walrath, R., Willis, J. O., Dumont, R., & Kaufman, A. S. (2019). Factor-Analytic Models of Intelligence. *En R.J. Sternberg (Ed). The Cambridge Handbook of Intelligence, (pp.75–98)*. Cambridge: Cambridge University Press. doi:10.1017/9781108770422.006
- Wang, W.C., & Shih, C.L. (2010). MIMIC Methods for Assessing Differential Item Functioning in Polytomous Items. *Applied Psychological Measurement, 34*(3), 166–180. doi:10.1177/0146621609355279
- Waschl, N., & Burns, N. R. (2020). Sex differences in inductive reasoning: A research synthesis using meta-analytic techniques. *Personality and Individual Differences, 164*, 109959. doi: 10.1016/j.paid.2020.109959
- Wichert, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist, 30*(7), 1006-1016. <https://doi.org/10.1080/13854046.2016.1205136>
- Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). A Cross-Temporal Meta-Analysis of Raven's Progressive Matrices: Age groups and developing versus developed countries. *Intelligence, 49*, 1-9. <https://doi.org/10.1016/j.intell.2014.11.008>
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*(5), 339-361. doi:10.1177/0146621611405984
- Zumbo, B. D. (1999). *A handbook on the theory and methods of Differential Item Functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense. Recuperado de: <http://www.faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf>