

XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional. II Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires, 2020.

Análisis psicométrico de la Escala de Compañerismo Universitario con la teoría de la respuesta al ítem.

Aune, Sofía, Galibert, María Silvia, Pralong, Héctor Omar y Attorresi, Horacio Felix.

Cita:

Aune, Sofía, Galibert, María Silvia, Pralong, Héctor Omar y Attorresi, Horacio Felix (2020). *Análisis psicométrico de la Escala de Compañerismo Universitario con la teoría de la respuesta al ítem*. XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional. II Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.

Dirección estable: <https://www.aacademica.org/000-007/114>

ARK: <https://n2t.net/ark:/13683/etdS/78a>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

ANÁLISIS PSICOMETRICO DE LA ESCALA DE COMPAÑERISMO UNIVERSITARIO CON LA TEORÍA DE LA RESPUESTA AL ÍTEM

Aune, Sofía; Galibert, María Silvia; Pralong, Héctor Omar; Attorresi, Horacio Felix
Universidad Buenos Aires. Facultad de Psicología. Buenos Aires, Argentina.

RESUMEN

El objetivo de la presente investigación fue explorar de las propiedades psicométricas de una Escala de Compañerismo Universitario (ECU) en construcción mediante la Teoría de la Respuesta al Ítem (TRI), con el fin de optimizar una versión definitiva. La muestra fue obtenida por accesibilidad y estuvo compuesta por estudiantes universitarios argentinos (82% mujeres). Posteriormente a verificar los supuestos de unidimensionalidad e independencia local se comparó el ajuste relativo de cuatro modelos de la TRI a los datos: el Modelo de Respuesta Graduada (MRG), el Modelo de Respuesta Graduada restringido, el Modelo Generalizado de Crédito Parcial y el Modelo de Crédito Parcial, determinándose que el MRG era el más adecuado para la modelización. Uno de los ítems presentó Funcionamiento Diferencial del Ítem por género no uniforme de grado severo y fue eliminado de la Escala. Los parámetros de umbral promedio tendieron a ubicarse en niveles bajos de compañerismo universitario, mientras que los parámetros de inclinación tuvieron valores altos. Esto último indica que las categorías de respuesta fueron aptas para distinguir entre sujetos con diferentes niveles de compañerismo universitario. El análisis con la Función de Información del Test mostró una precisión alta para niveles desde muy bajos hasta medios de rasgo.

Palabras clave

Compañerismo - TRI - DIF - Comparación

ABSTRACT

PSYCHOMETRIC ANALYSIS OF THE UNIVERSITY FELLOWSHIP SCALE WITH ITEM RESPONSE THEORY

The objective of the present investigation was to explore the psychometric properties of a University Fellowship Scale (UFS) under construction using the Item Response Theory (IRT), in order to optimize a definitive version. The sample was obtained by accessibility and was composed of Argentine university students (82% women). After verifying the assumptions of unidimensionality and local independence, the relative fit of four models of the IRT to the data was compared: Graded Response Model (GRM), Reduced Graded Response Model, Generalized Partial Credit Model and Partial Credit Model, determining

that the GRM was the most suitable for modeling. One of the items presented non-uniform gender Differential Item Functioning with severe degree and was eliminated from the Scale. The average threshold parameters tended to be located at low levels of university fellowship, while the slope parameters had high values. The latter indicates that the response categories were apt to distinguish between subjects with different levels of university fellowship. The analysis with the Test Information Function showed high precision for levels from very low to average of the trait.

Keywords

University fellowship - Item Response Theory - Comparison between models - Differential Item Function

El compañerismo es una conducta prosocial en tanto se trata de una acción positiva hacia otro (Auné, Blum, Abal, Lozzia, & Attorresi, 2014). Puede encuadrarse en diferentes ámbitos, aunque la mayoría de los estudios psicológicos y sociales que abordan el tema en el contexto educativo se centran en los niveles de enseñanza primario y secundario (e.g. Rodríguez, Resett, Grinóvero, & Moreno, 2015). Así, se estudia al compañerismo como parte del interés más general por las habilidades y los vínculos sociales de los alumnos, y como concepto contrario a los de *bullying* y agresión. En adultos, en cambio, se ha dado énfasis al compañerismo en el contexto laboral (e.g. Collado, Soria, Canafoglia, & Collado, 2016).

En el ámbito universitario, esta conducta sólo se ha abordado recientemente. Carranza Esteban, Hernández y Alhuay-Quispe (2017) hallaron una relación directa y estadísticamente significativa entre la dimensión de relaciones positivas de la escala adaptada de bienestar psicológico de Ryff, Lee, Essex, y Schmutte (1994) y el rendimiento académico de estudiantes universitarios. Concordantemente, Lara Grados y Prutsky Kalinkauský (2017) encontraron en alumnos de igual nivel educativo que a mayor desarrollo de habilidades sociales y trabajo en equipo, más importante será su influencia facilitadora en el aprendizaje colaborativo. Así, el rol social del estudiante cobra importancia como parte de la explicación del éxito académico. Se infiere que el compañerismo constituye un elemento fundamental para crear

un clima propicio que permita la incorporación de conocimientos. Se define en este trabajo al compañerismo universitario como el conjunto de acciones de ayuda y cooperación entre estudiantes de nivel superior que se realizan con el objetivo de beneficiar a un par en algún aspecto de su labor académica. Dada la falta de instrumentos de evaluación psicológica existentes para medir este constructo los autores han creado la Escala de Compañerismo Universitario (ECU). Se trata de una escala corta en construcción, cuyos ítems han sido redactados a partir de entrevistas a estudiantes y cuyas propiedades psicométricas se están explorando para poder optimizar una versión definitiva. Los ítems que componen actualmente la ECU son: 1) *Comparto con mis compañeros las experiencias que tuve en las materias que ya cursé*; 2) *Si mis compañeros están con dificultades, no me molesta hacer tareas de más en un trabajo de grupo*; 3) *Presto apuntes o libros a compañeros*; 4) *Explico a mis compañeros cómo hacer los trámites de la facultad* y 5) *Comparto mis conocimientos con compañeros*. Se ha diseñado una escala Likert de seis opciones para responder los ítems.

MÉTODO

Participantes

La muestra estuvo compuesta por 582 estudiantes argentinos de nivel superior (79% mujeres) que cursaban en la Universidad de Buenos Aires. La media de edad fue de 22.7 años con un desvío estándar de 6.1 (Mínimo = 18 años; Máximo = 60 años).

Procedimiento

Se recolectaron los datos mediante un diseño muestral no probabilístico por accesibilidad. El protocolo fue administrado en formato de lápiz y papel, donde se incluía un consentimiento informado. Además, se aclaraba que el uso era exclusivamente para fines de investigación y la participación enteramente voluntaria, pudiendo cesar en cualquier momento.

Instrumentos

Escala de Compañerismo Universitario (ECU). Se compone de cinco reactivos que se responden en una escala Likert de seis opciones (*Nunca, Casi Nunca, A Veces, Con Frecuencia, Casi Siempre, Siempre*). La ECU había sido, previamente, sometida a un proceso de depuración. Un banco de 20 ítems iniciales fue sometido a juicio experto en cuanto a congruencia con la definición de compañerismo universitario y relevancia. El acuerdo inter-jueces se analizó con el índice V de Aiken (Penfield y Giacobbi, 2004), conservándose los 5 ítems que obtuvieron mayor consenso. Posteriormente, se realizó un estudio piloto donde se administró la ECU a 35 estudiantes universitarios, con el fin de revisar que tanto la consigna como los ítems fueran claros y brindar evidencias de validez aparente.

Análisis de datos

Cumplimiento de los Supuestos de los Modelos de la TRI

La verificación del supuesto de unidimensionalidad requerido por los modelos MRG, MRG_R, MGCP y MCP a comparar se realizó mediante los índices UniCo (*Unidimensional Congruence*), ECV (*Explained Common Variance*) y MIREAL (*Mean of Item Residual Absolute Loadings*) obtenidos en un Análisis Factorial Exploratorio. Estos índices informan que los datos pueden ser tratados como esencialmente unidimensionales si: UniCo > .95; ECV > .85 y MIREAL < .30 (Ferrando & Lorenzo-Seva, 2017a).

Además del supuesto de unidimensionalidad, los modelos MRG, MRG_R, MGCP y MCP también asumen que, dado un nivel fijo de *theta*, las respuestas a los ítems son independientes entre sí. El índice χ^2_{LD} (Chen & Thissen, 1997) se calcula para cada par de ítems e indica el no cumplimiento del supuesto cuando es mayor que 10.

Comparación entre Modelos de la TRI

Para comparar el ajuste relativo de los modelos MRG, MRG_R, MGCP y MCP se utilizaron múltiples métodos, tal como lo describen De Ayala (2009) y Toland (2013). Por una parte, se implementó el Test de la Razón de Verosimilitudes (*Likelihood Ratio Test*; LRT) que compara dos modelos anidados, complementado con el estadístico (Haberman, 1978). En este caso el MRG_R es una restricción del MRG y el MCP del MGCP.

Análisis del Funcionamiento Diferencial del Ítem

Se exploró la existencia de DIF por género mediante la serie de pasos detallados por Woods (2009). Como primer paso, cada uno de los reactivos de la ECU se analizó con el test de Wald modificado (Cai, 2012; Cai, Thissen, & du Toit, 2011; Langer, 2008) considerando el resto como anclaje. Posteriormente, se realizó un segundo paso, donde se testeó a un ítem con potencial DIF con anclaje en las respuestas al ítem más certeramente libre de DIF, evitando así la contaminación. El tamaño del efecto del DIF del ítem candidato fue evaluado con el Índice de DIF No Compensatorio (*Non-Compensatory DIF [NCDIF] Index*) considerando que existe DIF severo si NCDIF = .009 (Meade, Lautenschlager, & Johnson, 2007).

Evidencias en el Marco de la TRI

Una vez depurados los ítems por DIF, se obtuvieron evidencias de confiabilidad en el marco de la TRI, tanto para el ítem como para el test. La FII indica la precisión de determinado ítem en la medición de cada nivel de *theta*. La sumatoria de las FII conforma la FIT, que informa la confiabilidad de la escala según el nivel en el rasgo *theta*. La Confiabilidad Marginal permite conocer la fiabilidad global de la escala. Estas medidas fueron comparadas con el índice de confiabilidad clásico alfa de Cronbach.

RESULTADOS

Cumplimiento de los Supuestos de los Modelos de la TRI

Los índices de unidimensionalidad esencial tuvieron los siguientes valores: UniCo = .986, ECV = .879 y MIREAL = .215. Por lo tanto, los datos pueden considerarse esencialmente unidimensionales. En las salidas de cada uno de los modelos MRG, MRG_R, MGCP y MCP el índice X^2_{LD} resultó menor que 10 para cada par de ítems. Por lo tanto, para los cuatro modelos se pueden dar por satisfechos los dos supuestos, unidimensionalidad e independencia local.

Comparación entre Modelos de la TRI

Los índices de ajuste para MRG, MRG_R, MGCP y MCP se pueden observar en la Tabla 1. Aunque todos los modelos ajustaron globalmente, el MRG obtuvo valores más pequeños en estadístico M_2 , logaritmo de la verosimilitud, AIC y BIC. El LRT entre el MRG y el MRG_R indicó que la complejidad adicional del modelo com-

pleto es necesaria para mejorar el ajuste a los datos ya que $(4) = 8432.83 - 8368.99 = 63.84, p = 4.51 \times 10^{-13}$. El cambio relativo entre estos modelos fue de = .0075, es decir que el MRG mejora la explicación de los datos por sobre el MRG_R en un 0.75%. Muy similares resultados se obtienen en la comparación del MGCP y el MCP. El LRT entre el MGCP y el MCP resultó de $(4) = 8472.41 - 8407.11 = 65.30, p = 2.22 \times 10^{-13}$. En este caso el = .0077, señalando una mejora en el modelo completo de 0.77%. Consistentemente, en los modelos sin restricciones en el parámetro a MRG y MGCP, ningún ítem desajustó ya que el p asociado al S- X^2 fue mayor que .01, para todos los reactivos. En cambio, dos ítems no ajustaron al MRG_R y tres al MCP.

Dado que un modelo con parámetro a libre es necesario para mejorar tanto el ajuste como la explicación de estos datos y que el MRG es el modelo con mejor ajuste relativo, se lo selecciona para la modelización con TRI de las respuestas a los ítems de la ECU.

Tabla 1
Comparación de Ajuste a la ECU entre Modelos de la TRI

Modelo	M_2	g/l	p valor	RMSEA	-2lnL	AIC	BIC	Desaj
MRG	389.87	245	.0001	.03	8368.99	8428.99	8559.98	No
MRG _R	481.56	249	.0001	.04	8432.83	8484.83	8598.36	Sí
MGCP	534.85	245	.0001	.05	8407.11	8467.11	8598.10	No
MCP	634.84	249	.0001	.05	8472.41	8524.41	8637.94	Sí

Nota: ECU = Escala de Compañerismo Universitario; TRI = Teoría de la Respuesta al Ítem; M_2 = estadístico M_2 ; g/l = grados de libertad; RMSEA = error medio cuadrático de aproximación; -2lnL = valor del logaritmo de la verosimilitud; AIC = Criterio de Información de Akaike; BIC = Criterio de Información bayesiano; Desaj = Existencia de desajuste de ítems al modelo según el índice S- γ^2 ; MRG = Modelo de Respuesta Graduada; MRG_R = MRG con restricción de igualdad en el parámetro a ; MGCP = Modelo Generalizado de Crédito Parcial; MCP = Modelo de Crédito Parcial.

Modelización de la ECU con el MRG

El MRG fue, de los cuatro modelos, el que resultó más adecuado. Supone que un único γ -en este caso el nivel de compañerismo universitario- explica no linealmente las respuestas a los ítems. Para cada uno de los reactivos fueron calculados un parámetro de inclinación a y cinco parámetros de umbral de categoría b_m , dado que las opciones de respuesta son seis. El parámetro a informa el grado en que las categorías de respuesta distinguen entre los niveles de γ . Además, se lo ha comparado con las cargas factoriales, en tanto refleja la magnitud de la relación de cada ítem de la escala con el rasgo latente γ . Cada parámetro b_m informa el nivel de compañerismo universitario γ - que se necesita para tener igual probabilidad (.50) de seleccionar la categoría de respuesta m o superior con respecto a las categorías inferiores. Así, las opciones de respuesta de los ítems son separadas en una serie de dicotomías en cada una de las cuales se aplica el ML2P. Las CCCRI representan la probabilidad de seleccionar cada categoría de respuesta según el nivel de γ .

Análisis del Funcionamiento Diferencial de los Ítems

Los resultados de explorar la presencia de DIF utilizando para cada ítem al resto como anclaje se encuentran en la Tabla 2. Los parámetros de los ítems por cada género estimados con el MRG se observan en la Tabla 3. El ítem 2 - *Si mis compañeros están con dificultades, no me molesta hacer tareas de más en un trabajo de grupo* - fue señalado por los análisis como candidato a tener DIF no uniforme por género. Su contenido parece no relacionarse con el compañerismo universitario para los varones ($a_{\text{varones}} = 0.30$) pero sí en el caso de las mujeres ($a_{\text{mujeres}} = 0.94$). Sus parámetros b_m , demasiado alejados del rango esperable -3 a 3, indican un claro desajuste del ítem para los varones. Se implementó el segundo paso del método de Woods (2009), tomando al ítem 5 como anclaje entre los dos grupos. Este ítem tiene el menor valor de χ^2_{Total} y además los parámetros a y b_m son iguales para los dos grupos, por lo cual se asume libre de DIF. El resultado estadísticamente significativo corroboró que el ítem 2 exhibe DIF ($\chi^2_a = 8.7, g/l = 1, p = .0032$). En cuanto al tamaño del efecto, se obtuvo un índice NCDIF de .1612, es decir que se trata de un DIF severo. Por lo tanto, el ítem 2 será eliminado de la ECU y de los posteriores análisis de este trabajo.

Tabla 2
Análisis de Funcionamiento Diferencial del Ítem entre Mujeres y Varones. χ^2 de Wald

Ítem	χ^2_{Total} (gl 6)	<i>p</i>	χ^2_a (gl 1)	<i>p</i>	χ^2_b (gl 5)	<i>p</i>
1	4.8	0.5698	0.5	0.4628	4.3	0.5128
2	13.3	0.0390*	7.7	0.0056*	5.6	0.3517
3	3.1	0.7954	0.1	0.7520	3.0	0.6991
4	4.9	0.5589	0.0	0.8650	4.9	0.4342
5	2.1	0.9083	0.4	0.5485	1.8	0.8813

Nota. El χ^2_{Total} se refiere al test global, el χ^2_a al test para DIF no uniforme y el χ^2_b al test para DIF uniforme. *Se consideran significativos los $p < .05$.

Tabla 3

Parámetros de los ítems estimados por MRG para el análisis de DIF por género

Ítem	Gen	<i>a</i> (s.e.)	<i>b</i> ₁ (s.e.)	<i>b</i> ₂ (s.e.)	<i>b</i> ₃ (s.e.)	<i>b</i> ₄ (s.e.)	<i>b</i> ₅ (s.e.)
1	M	2.08(0.26)	-3.33(0.38)	-2.74(0.26)	-1.58(0.13)	-0.75(0.09)	0.06(0.08)
	V	1.54(0.57)	-3.77(1.36)	-2.76(0.95)	-1.64(0.56)	-0.52(0.25)	0.56(0.32)
2*	M	0.94(0.12)	-4.11(0.53)	-2.19(0.27)	-0.01(0.11)	0.95(0.16)	2.39(0.29)
	V	0.30(0.18)	-10.79(6.59)	-5.46(3.29)	-0.74(0.74)	2.52(1.72)	6.36(3.97)
3	M	1.38(0.15)	-3.62(0.41)	-2.00(0.19)	-0.76(0.10)	0.50(0.10)	1.44(0.15)
	V	1.34(0.56)	-3.44(1.32)	-1.86(0.69)	-0.77(0.32)	0.36(0.29)	1.46(0.64)
4	M	2.18(0.26)	-2.40(0.20)	-1.87(0.15)	-0.77(0.08)	0.12(0.07)	0.76(0.09)
	V	2.06(0.92)	-3.01(1.13)	-1.80(0.64)	-0.67(0.27)	0.11(0.21)	0.94(0.44)
5	M	1.31(0.17)	-5.08(0.84)	-3.08(0.38)	-1.30(0.16)	-0.16(0.09)	0.93(0.15)
	V	1.31(0.17)	-5.08(0.84)	-3.08(0.38)	-1.30(0.16)	-0.16(0.09)	0.93(0.15)

Nota. Gen: género, M: mujer, V: varón; *a*: parámetro de inclinación calculado con el MRG; *b*₁, *b*₂, *b*₃, *b*₄ y *b*₅: parámetros de umbral calculados con el MRG. *Ítem que exhibe DIF no uniforme

Calibración de los ítems con el MRG

Se realizó una nueva corrida del MRG excluyendo al ítem 2. Los resultados indicaron que el modelo ajustó tanto globalmente ($M_2 = 248.61$; $gl = 146$; $p = 0.0001$; $RMSEA = 0.03$) como al nivel de los ítems (p asociado al $S-X^2 > .01$). Se estimaron 24 parámetros, cuyos valores se exponen en la Tabla 4.

Los parámetros *b*₁ fueron de a lo sumo -2.43, es decir que para los cuatro ítems es suficiente un nivel muy bajo de compañerismo universitario para tener igual probabilidad de elegir *Nunca* versus la categoría *Casi Nunca* o una superior en la escala Likert. Los parámetros *b*₂ tuvieron valores de como máximo -1.79, indicando que para tener una probabilidad de 0.5 de seleccionar la categoría *A Veces* (o una superior) sobre las categorías previas es suficiente, en todos los ítems, con tener un nivel bajo de compañerismo universitario. Aunque distante de los anteriores, el parámetro *b*₃ sigue ubicado en los niveles bajos del rasgo.

Es el parámetro *b*₄, es decir el nivel de rasgo necesario para seleccionar la categoría *Casi Siempre* o *Siempre* por sobre las anteriores, el primero que requiere para dos de los ítems niveles de compañerismo universitario superiores al promedio. Por último, el parámetro *b*₅, nivel de compañerismo universitario necesario para tener igual probabilidad de seleccionar la categoría *Siempre* con respecto a todas las anteriores, se ubicó para los cuatro ítems en niveles por encima de la media, aunque con variaciones importantes entre los reactivos.

Los parámetros de inclinación *a* fueron de como mínimo 1.38. Esto indica que los reactivos se vinculan fuertemente con compañerismo universitario y sus categorías de respuesta poseen una capacidad elevada para diferenciar entre sujetos con diferentes niveles de *theta* (Reise & Waller, 1990). Las CCCRI de cada uno de los ítems se pueden observar en las Figuras 1, 2, 3 y 4.

Tabla 4
Parámetros estimados mediante MRG para los ítems de la ECU

Ítem	$\alpha(s.e.)$	$b_1(s.e.)$	$b_2(s.e.)$	$b_3(s.e.)$	$b_4(s.e.)$	$b_5(s.e.)$	$b_{promedio}$
1	2.05(0.23)	-3.45(0.31)	-2.59(0.21)	-1.51(0.12)	-0.67(0.07)	0.15(0.07)	-1.61
3	1.38(0.14)	-3.48(0.34)	-1.91(0.17)	-0.72(0.09)	0.48(0.09)	1.43(0.13)	-0.84
4	2.19(0.24)	-2.43(0.19)	-1.79(0.13)	-0.72(0.07)	0.14(0.06)	0.78(0.08)	-0.80
5	1.40(0.14)	-4.75(0.63)	-2.89(0.26)	-1.22(0.12)	-0.14(0.08)	0.89(0.11)	-1.62

Nota. MRG: Modelo de Respuesta Graduada; ECU: Escala de Compañerismo Universitario; α : parámetro de inclinación calculado con el MRG; b_1 , b_2 , b_3 , b_4 y b_5 : parámetros de umbral calculados con el MRG. *s.e.*: error estándar; $b_{promedio}$: promedio de los parámetros de umbral b .

Figura 1
Curvas Características de las Categorías de Respuesta del Ítem 1

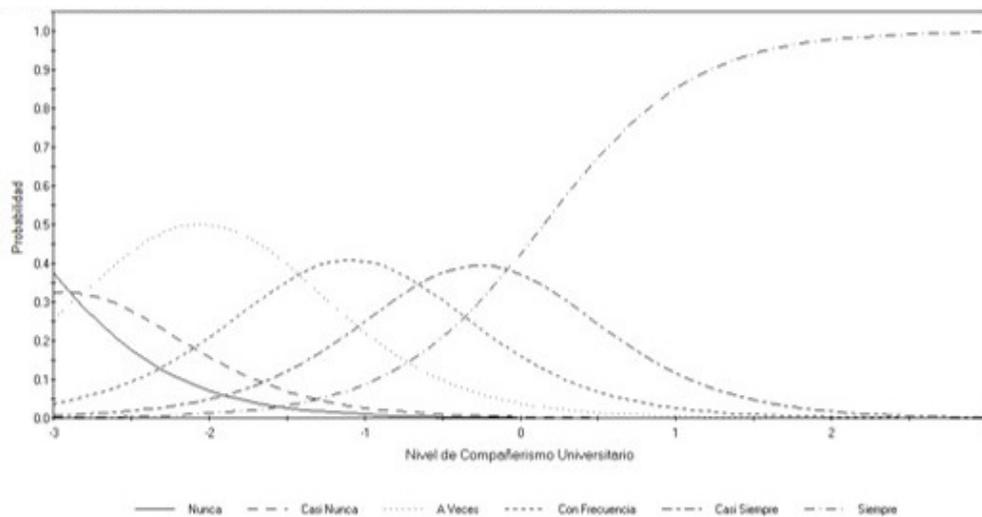


Figura 2
Curvas Características de las Categorías de Respuesta del Ítem 3

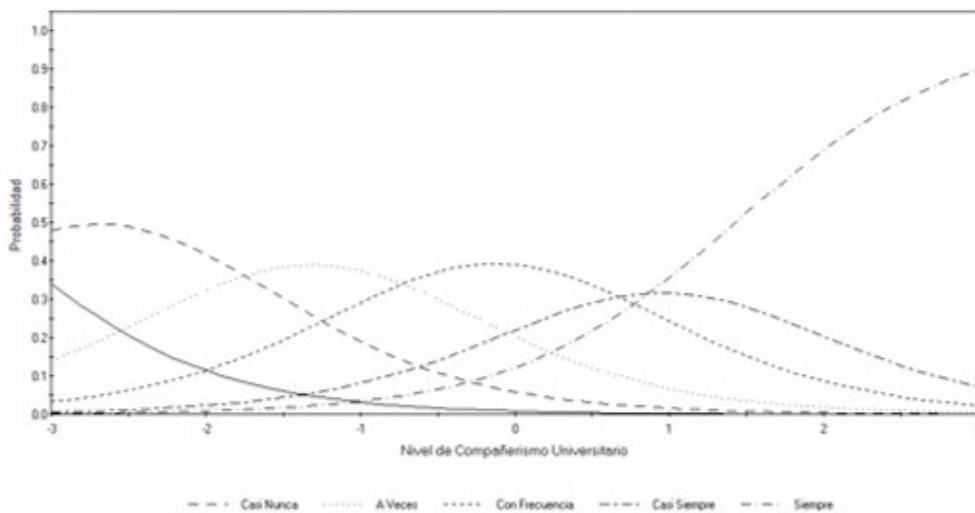


Figura 3
Curvas Características de las Categorías de Respuesta del Ítem 4

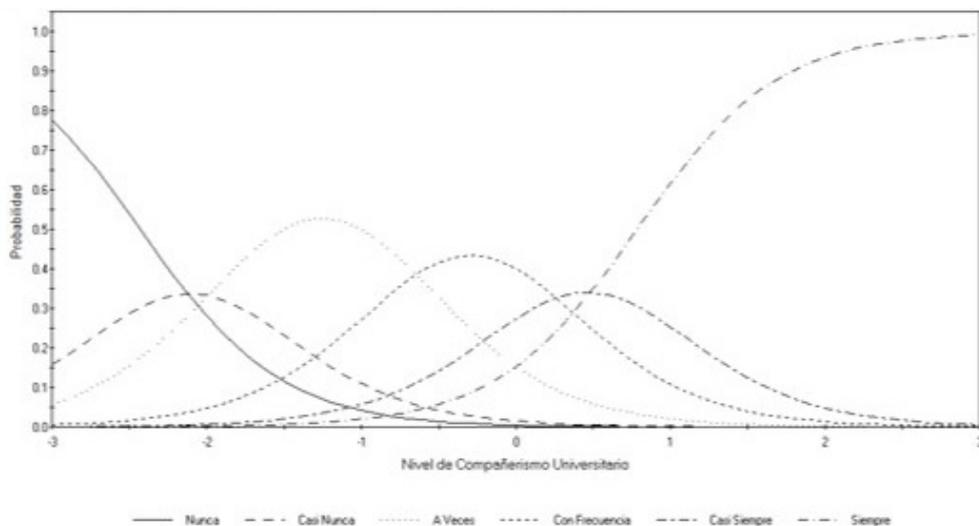
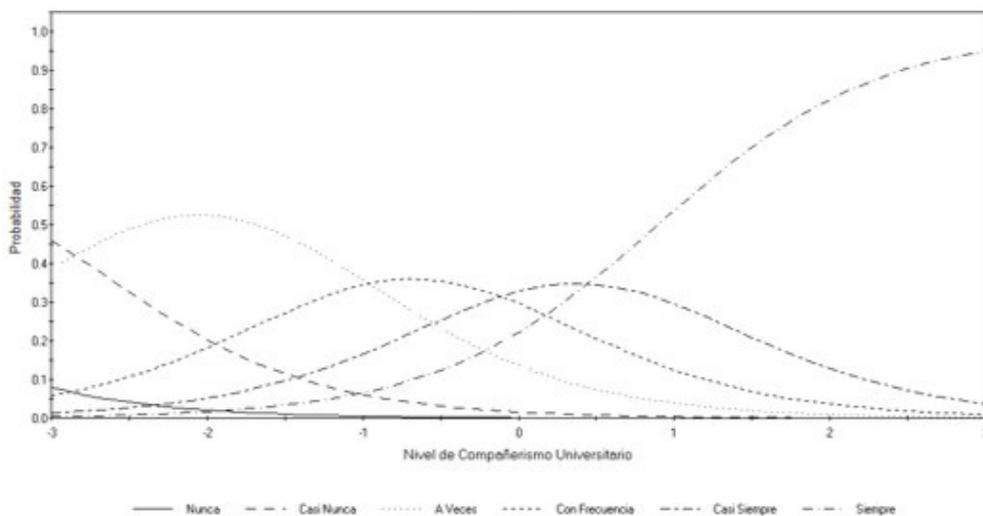


Figura 4
Curvas Características de las Categorías de Respuesta del Ítem 5

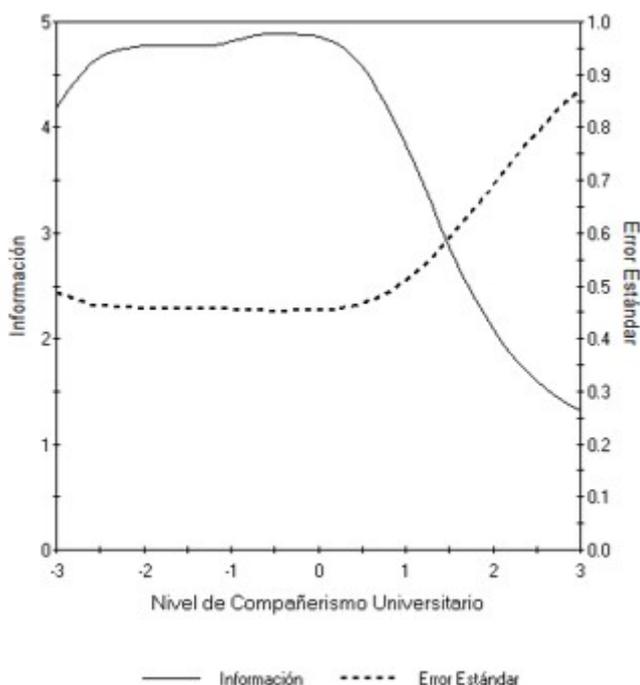


Evidencias de Confiabilidad

La Figura 5 muestra la FIT de la ECU estimada con el MRG. La confiabilidad fue elevada para sujetos que se ubican desde tres desvíos estándar por debajo de la media de compañerismo universitario hasta un desvío estándar por sobre la media formando

así una meseta. A medida que el nivel de rasgo aumenta, cae el nivel de información y sube el error estándar de medición. El valor del índice de confiabilidad marginal fue de .76, muy adecuado por tratarse de una prueba tan corta. El alfa de Cronbach fue de .73, también aceptable en este caso.

Figura 5
Función de Información del Test



DISCUSIÓN

El análisis de la ECU con el MRG evidenció que la escala tiene un nivel más elevado de discriminación en niveles medios y bajos del rasgo. El error de medición crece sustancialmente hacia niveles altos de compañerismo universitario. La capacidad discriminativa, así como el nivel de información alcanzado tuvieron valores adecuados para la totalidad de los ítems que componen la ECU. Los nuevos ítems que se incorporen a la ECU deberían requerir un nivel alto de compañerismo universitario para que la escala definitiva pueda medir con similar precisión a los distintos niveles del rasgo.

En cuanto al análisis de la adecuación de la cantidad de opciones de respuesta, valores demasiado bajos en los parámetros b_1 implican que podría ser preferible una reducción a cinco categorías. También podría ser de interés testear la respuesta con cuatro categorías para evitar la tendencia a elegir la opción central. Además, resultados empíricos y en base a simulación indican que un diseño de respuesta con cuatro opciones favorece el equilibrio entre la precisión de la medida y el grado de ajuste del modelo de la TRI (e.g. Abal, Auné, Lozzia & Attorresi, 2017; Lozano, García-Cueto & Muñiz, 2008).

Con respecto a un análisis detallado de los reactivos, los ítems 1 y 4 fueron los que brindaron los niveles más altos de información. Además, el parámetro de inclinación a de estos ítems es muy alto y las distancias entre los parámetros b_m son amplias, lo que indica que las categorías de respuesta resultan eficaces para discriminar entre participantes con diferentes niveles de

compañerismo. Los ítems 3 y 5, si bien tienen una calidad psicométrica aceptable, son menos informativos que los anteriores. El análisis del DIF tuvo como resultado la detección de un ítem que presentaba funcionamiento diferencial no uniforme por género. Esto muestra la importancia de los estudios del DIF, tan poco frecuentes sobre todo en el medio latinoamericano. La existencia de ítems con DIF resta validez a la interpretación de los puntajes de la escala, de manera que es necesario testear todos los ítems que se incorporen a la ECU en este sentido. Además del DIF por género, se puede analizar con respecto a variables sociodemográficas e incluso psicológicas obteniendo así resultados de mucho interés. En futuros estudios se analizará el DIF por edad y por nivel del participante en la Escala de Conducta Empática (Auné et al., 2017a) y en la adaptación argentina de la Escala de Felicidad de Lima (Auné, Abal, & Attorresi, 2017b). Otra cuestión para destacar es la relevancia de la comparación entre modelos de la TRI. Se pudo observar que no todos los modelos de la TRI ajustaron en el mismo grado a los datos obtenidos empíricamente, las respuestas a la ECU de esta muestra en particular. Los modelos aplican distintas formas de segmentación del ítem politómico y utilizan diferentes procedimientos para el cálculo de las probabilidades de respuesta de las categorías. Para los tests de comportamiento típico, los resultados de este estudio concuerdan con otros donde el MRG ajustó mejor que los restantes modelos comparados (e.g. Abal, 2013; Asún & Zuñiga, 2008).

REFERENCIAS

- Abal, F. (2013). *Comparación de modelos politómicos y dicotómicos de la Teoría de la Respuesta al Ítem aplicados a un test de Comportamiento Típico*. Tesis de Doctorado, Facultad de Psicología de la Universidad de Buenos Aires.
- Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2017). Funcionamiento de la categoría central en ítems de Confianza para la Matemática. *Evaluar*, 17(2), 18-31.
- Asún, R. & Zuñiga, C. (2008). Ventaja de los modelos politómicos de teoría de respuesta al ítem en la medición de actitudes sociales. El análisis de un caso. *Psyche*, 17(2), 103-115.
- Auné, S., Abal, F., & Attorresi, H. (2017a). Propiedades psicométricas de una prueba de conducta empática. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 3(45), 47-56. <https://doi.org/10.21865/RIDEP45.3.04>
- Auné, S., Abal, F., & Attorresi, H. (2017b). Versión argentina de la Escala de Felicidad de Lima. *Diversitas*, 13(2), 201-214.
- Auné, S., Blum, G., Abal, F., Lozzia, G., & Attorresi, H. (2014). La Conducta Prosocial: estado actual de la investigación. *Perspectivas en Psicología*, 11(2), 21-33.
- Cai, L. (2012). *flexMIRT: Flexible multilevel item factor analysis and test scoring* [Computer software]. Seattle, WA: Vector Psychometric Group, LLC.
- Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO user's guide*. Lincolnwood, IL: Scienti?c Software International.

- Carranza Esteban, R. F., Hernández, R. M., & Alhuay-Quispe, J. (2017). Bienestar psicológico y rendimiento académico en estudiantes de pregrado de psicología. *Revista Internacional de Investigación en Ciencias Sociales*, 13(2), 133-146.
- Chen, W., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289.
- Collado, P. A., Soria, C. B., Canafoglia, E., & Collado, S. A. (2016). Condiciones de trabajo y salud en docentes universitarios y de enseñanza media de Mendoza, Argentina: Entre el compromiso y el desgaste emocional. *Salud Colectiva*, 12, 203-220.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- Haberman, S. J. (1978). *Analysis of qualitative data: Vol. 1: Introductory topics*. New York, NY: Academic Press.
- Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.
- Lara Grados, C. E., & Prutsky Kalinkauský, J. (2017). Desarrollo de habilidades sociales, trabajo en equipo y nivel de aprendizaje colaborativo en estudiantes de la Facultad de Psicología de la Universidad Inca Garcilaso de la Vega. *Revista Peruana de Psicología y Trabajo Social*, 6(1), 63-84.
- Lozano, L. M., García-Cueto, E. & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73-79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Meade, A. W., Lautenschlager, G. J., & Johnson, E. C. (2007). A Monte Carlo examination of the sensitivity of the differential functioning of items and tests framework for tests of measurement invariance with Likert data. *Applied Psychological Measurement*, 31, 430-455. <https://doi.org/10.1177/0146621606297316>
- Penfield, R. D. y Giacobbi, P. R. (2004). Applying a Score Confidence Interval to Aiken's Item Content-Relevance Index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/s15327841mpee0804_3
- Rodríguez, L., Resett, S., Grinóvero, M., & Moreno, J. (2015). Propiedades psicométricas de la Escala de Calidad de la Amistad en español. *Anuario de Psicología*, 45(2), 219-234.
- Ryff, C., Lee, Y. H., Essex, M. J., & Schmutte, P. S. (1994). My children and me: Midlife evaluations of grown children and of self. *Psychology and Aging*, 9(2), 195-205.
- Toland, M. (2013). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, 34(1), 120-151. <https://doi.org/10.1177/0272431613511332>
- Woods, C. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57. <https://doi.org/10.1177/0146621607314044>