

## Funcionamiento de la Categoría Central en Ítems de Confianza para la Matemática

### The Role of a Central Category among Items in a Confidence in Mathematics Scale

Facundo Juan Pablo Abal \* <sup>1,2</sup>, Sofía Esmeralda Auné <sup>1,2</sup>, Gabriela Susana Lozzia <sup>2</sup>,  
Horacio Félix Attorresi <sup>2</sup>

1 - Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

2 - Universidad de Buenos Aires, Argentina.

Introducción  
Método  
Resultados  
Discusión  
Referencias

Recibido: 28/04/2017 Revisado: 30/05/2017 Aceptado: 01/06/2017

#### Resumen

Se comparan las propiedades psicométricas observadas en un test de Confianza para la Matemática al utilizar formatos de respuesta Likert con y sin categoría central. El instrumento mide un conjunto de creencias del estudiante sobre sus dificultades para responder a las habilidades que demanda la matemática. En el estudio participaron 939 estudiantes de psicología (81% mujeres), los cuales completaron el instrumento con 2 formatos Likert de: a) 5 opciones con categoría intermedia *Ni de acuerdo ni en desacuerdo* y b) 6 opciones con dos categorías centrales (*Más bien en desacuerdo* y *Más bien de acuerdo*). La variación de la escala Likert no afectó sustancialmente las evidencias de validez basadas en la estructura interna (análisis factorial confirmatorio y ajuste al Modelo de Crédito Parcial) ni la relación con otras variables. La función de eficiencia relativa reveló que se obtiene similar información para todos los niveles del rasgo.

**Palabras clave:** *escala Likert, categoría central, Confianza para la Matemática, estudiantes universitarios*

#### Abstract

This study compares the psychometric properties of a Confidence for Mathematics Scale using Likert response formats with and without a central category. The scale measures a set of student beliefs about their difficulties to respond to the skills required in the field of Mathematics. The participants were 939 Psychology students (81% female), who responded to the items using two response formats: a) a 5-point scale with the central category *Neither agree nor disagree* and b) a 6-point scale with 2 central categories: *Slightly Disagree* and *Slightly Agree*. Variations on the Likert scale did not substantially affect the evidence of validity based on internal structure (confirmatory factor analysis and Partial Credit Model fit) or in relations with other variables. The relative efficiency function revealed that similar information is obtained for all levels of the trait.

**Keywords:** *Likert scale, central category, Confidence in Mathematics, college students*

\*Correspondencia a: Facundo Juan Pablo Abal. Dirección postal: Zuviría 5691 (1439), Ciudad Autónoma de Buenos Aires. Teléfono: 54 11 46014331.

[fabal@psi.uba.ar](mailto:fabal@psi.uba.ar)

**Cómo citar este artículo:** 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. *Revista Evaluar*, 17(2), 18-31. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

#### Financiación

Esta investigación fue financiada con subsidios de la Universidad de Buenos Aires (UBACyT 2014-17, Código 20020130100320BA) y de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 2011 N° 0826).

## Introducción

### *Funcionamiento de la categoría central en ítems de Confianza para la Matemática*

El debate acerca de la conveniencia de incluir o no una categoría central en las escalas Likert ha generado una multiplicidad de opiniones y abordajes metodológicos (Bisquerra & Pérez-Escoda, 2015; Joshi, Kale, Chandel, & Pal, 2015; Tsang, 2012). La recomendación habitual es que la decisión quede supeditada al juicio prudencial del elaborador de la prueba (Morales, 2006). No obstante, en la gran mayoría de los casos se suele escoger un formato Likert con categoría central (principalmente con cinco opciones) justificado más en la tradición y el pragmatismo que en fundamentos psicométricos (Muñiz, García-Cueto, & Lozano, 2005).

Las posturas a favor de su inclusión sostienen que los evaluados tienden a sentirse más a gusto cuando no se ven compelidos a adoptar una decisión frente a temas para los que se encuentran genuinamente indecisos. Además, la incorporación de categorías más sutiles permitiría discriminar mejor sus respuestas, con el aumento consecuente de la detección de diferencias individuales. Esto último ha sido comprobado especialmente con la incorporación de una categoría intermedia en ítems dicotómicos (Morales, 2006). Los detractores de una respuesta intermedia consideran que su inclusión perjudica las propiedades psicométricas del instrumento, dado que su elección podría estar influenciada por la deseabilidad social (Johns, 2005), ciertas características del enunciado del ítem como ambigüedad y descontextualización (Kulas & Stachowski, 2013) y rasgos de personalidad (Murray, Booth, & Molenaar, 2016).

Uno de los principales argumentos en contra de la categoría intermedia cuestiona la presunción de que los evaluados interpretan de forma inva-

riable los cuantificadores lingüísticos que generalmente acompañan a esta categoría de respuesta (Hernández, Drasgow, & González-Romá, 2004). Idealmente, los evaluados responden a la categoría intermedia cuando su postura legítima está entre las opciones de desacuerdo y acuerdo en el continuo de la escala de respuesta. De esta manera sus respuestas reflejarían niveles moderados del constructo. Sin embargo, Kulas y Stachowski (2009) mostraron que la categoría central también es elegida por los evaluados cuando estos no están seguros de sus opiniones, pretenden ocultarlas o no comprenden cabalmente el enunciado del ítem.

En los últimos años, los modelos politómicos de la Teoría de Respuesta al Ítem (TRI) han tomado protagonismo en el análisis del efecto del formato Likert sobre la validez y confiabilidad de los tests, a través de estudios empíricos y con datos simulados (p. ej. Lee & Paek, 2014; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol, & Coffman, 2009). Más específicamente, los estudios se centraron en analizar los ítems con modelos politómicos que, durante el proceso de estimación, no condicionan un orden para los parámetros correspondientes a las categorías de respuesta de los ítems. Tal es el caso de los modelos de respuesta nominal (Thissen & Cai, 2016) o de crédito parcial (Masters, 2016). De esta manera, en los mencionados estudios se intenta corroborar si efectivamente la categoría intermedia resulta eficaz para captar valores centrales del rasgo (Andrich, de Jong, & Sheridan, 1997; González & Espejo, 2003; Hernández, Espejo, & González, 2006; Hernández, Espejo, González, & Gómez, 2001; Murray et al., 2016; Preston, Reise, Cai, & Hays, 2011).

En el presente trabajo se propone analizar el efecto que produce la inclusión de una categoría central en la escala Likert sobre las propiedades psicométricas de un test. Por tratarse de un estudio empírico, se plantea la comparación de

los resultados obtenidos en una prueba que mide la *confianza para la Matemática* en estudiantes de psicología (Abal, Auné, & Attorresi, 2014). Este constructo permite examinar un conjunto de creencias del estudiante sobre sus posibilidades y dificultades para responder a las habilidades requeridas en la actividad matemática. Resulta particularmente relevante en el contexto de carreras universitarias de corte humanístico-social en las cuales los estudiantes tienden a desconfiar de sus capacidades y perjudican así su implicación en el proceso de aprendizaje (Blanco, 2004; Narro, 1997). Las creencias vinculadas a la propia capacidad percibida por el alumno para enfrentar situaciones de manera exitosa presentan una fuerte carga afectiva y juegan un papel esencial en la concreción de logros (Bermejo, 1996; Gil, Blanco, & Guerrero, 2006; Gómez-Chacón, 2005; Gómez-Chacón, Op't Eynde, & De Corte, 2006; McLeod & McLeod, 2002). En estudios sobre las actitudes vinculadas a la matemática se ha incluido sistemáticamente un componente destinado a la medición de la confianza para la matemática recibiendo estas denominaciones tales como Actitud hacia el éxito en matemática (Fennema & Sherman, 1976), Autopercepción (Adelson & McCoach, 2011), Habilidad (Bazán & Sotero, 1998), Autoconcepto matemático (Palacios, Arias, & Arias, 2014) y Confianza (Auzmendi, 1992; Tapia & Marsh, 2004).

Con base en estas consideraciones teóricas y metodológicas y con el fin de analizar el efecto de la inclusión de una categoría central, nuestro objetivo fue comparar las propiedades psicométricas verificadas en la Escala de Confianza para la Matemática (Abal et al., 2014) empleando ítems con formatos de respuesta de cinco y seis opciones.

A modo de objetivos específicos se planteó considerar: a) las variaciones en los valores adoptados en la estimación de parámetros del Modelo

de Crédito Parcial (MCP) de la TRI y en el ajuste de los datos a este modelo, b) las evidencias de validez basadas en la relación con otras variables externas y c) los estudios de confiabilidad clásicos y con TRI.

## Método

### *Participantes*

Los participantes que colaboraron en este estudio fueron 939 estudiantes de la carrera de Psicología (81% mujeres y 19% varones) residentes en el área metropolitana de Buenos Aires, Argentina. La edad osciló entre 18 y 59 años, con un valor promedio de 22.2 años ( $DE = 5.4$ ). El 25% debió recurrir al menos una vez la materia Matemática –correspondiente al primer año de la carrera– luego de haber obtenido un aplazo. La calificación promedio de los participantes en esta materia fue de 6.75 ( $DE = 1.84$ ).

### *Instrumentos*

**Cuestionario de variables sociodemográficas y académicas.** Este cuestionario recaba información acerca de características tales como género, edad y lugar de residencia. Asimismo, indaga sobre las condiciones de aprobación de Matemática: calificación obtenida y cantidad de veces que debió recurrir la materia antes de poder aprobarla.

**Escala de Confianza para la Matemática.** Esta prueba se compone de ocho ítems redactados en sentido inverso (Figura 1). El formato de respuesta original presenta una escala Likert con seis categorías: *Totalmente en desacuerdo*, *En desacuerdo*, *Más bien en desacuerdo*, *Más bien de acuerdo*, *De acuerdo* y *Totalmente de acuerdo*. Los alumnos con niveles elevados de confianza tienden a percibirse seguros y eficaces ante situaciones en

las que deben aplicar la Matemática. En cambio, los estudiantes con niveles bajos de confianza se reconocen con dificultades para incorporar, recuperar y aplicar lo aprendido. Estudios previos han aportado evidencias sobre validez basadas en la estructura interna: se verificó la unidimensionalidad del constructo mediante un análisis factorial exploratorio y se ajustó satisfactoriamente el Modelo de Respuesta Graduada de la TRI. En cuanto a la confiabilidad, se obtuvo un coeficiente alfa de .90 y un coeficiente de confiabilidad marginal de la TRI de .91 (Abal et al., 2014).

**Escala de Afecto hacia la Matemática.** Este instrumento mide el interés del estudiante en involucrarse en actividades vinculadas a la Matemática y los sentimientos asociados al uso de sus términos, símbolos y conceptos. La escala cuenta con evidencias de validez factorial obtenidas a partir de un análisis exploratorio y sus ítems han mostrado un ajuste aceptable al Modelo de Crédito Parcial. El análisis de la consistencia interna arrojó un índice alfa de Cronbach de .91 (Abal, Auné, Lozzia, & Attorresi, 2015).

**Escala de Utilidad de la Matemática.** Esta prueba evalúa un conjunto de creencias con respecto a la aplicabilidad de las nociones matemáticas en la formación académica y el futuro desempeño profesional. La escala fue validada para una población de estudiantes de Psicología y presenta adecuadas propiedades psicométricas. Todos los ítems se ajustaron al Modelo de Respuesta Graduada y la consistencia interna de la prueba medida mediante alfa de Cronbach fue de .89 (Abal, Galibert, Aguerri, & Attorresi, 2014).

#### *Procedimiento de recolección de datos*

Se adoptó un procedimiento que se utiliza habitualmente en estudios empíricos con objetivos similares (p. ej. González & Espejo, 2003;

Maydeu-Olivares et al., 2009). Los participantes respondieron un protocolo que contenía los ítems de confianza con dos formatos de respuesta: el formato original con seis categorías, y un formato de cinco opciones que resumía las dos opciones centrales con el anclaje *Ni de acuerdo ni en desacuerdo*. Otros autores han considerado la administración de un único formato de respuesta y un posterior reagrupamiento de las categorías en la etapa de análisis de datos (p. ej. Matell & Jacoby, 1971; Nunes et al., 2008). Sin embargo, consideramos que la estrategia implementada en este estudio permite analizar la reacción de los individuos frente a los anclajes lingüísticos usados en las escalas Likert.

Los estudiantes completaron el protocolo autoadministrado en grupos de aproximadamente 30 personas, coordinados por miembros del equipo de investigación. Durante la aplicación se aleatorizó el orden en que cada individuo respondió a los formatos, y se intercalaron los demás instrumentos administrados a fin de reducir el impacto de la memorización de las respuestas. La participación fue voluntaria, y se dieron garantías del anonimato y la confidencialidad de las respuestas. También se aclaró que la colaboración no tendría consecuencias negativas ni positivas con respecto al rendimiento académico, y que podían cesar de responder en cualquier punto de la evaluación.

#### *Análisis de datos*

**Unidimensionalidad.** A fin de verificar el supuesto de unidimensionalidad requerido por el MCP se implementó un análisis factorial confirmatorio (AFC) con el programa Mplus (Muthén & Muthén, 2010). Respetando el carácter ordinal de los datos, los parámetros fueron estimados con el método robusto de mínimos cuadrados ponderados (*Weighted Least Squares Mean and Variance*

*Adjusted*, WLSMV) sobre la base de las matrices de correlaciones policóricas. Se analizó el ajuste al modelo considerando los índices de ajuste comparativo (CFI) y de Tucker-Lewis (TLI), la raíz del promedio de las correlaciones residuales al cuadrado (RMSR) y el error medio cuadrático de aproximación (RMSEA). Siguiendo la recomendación de [Abad, Olea, Ponsoda y García \(2011\)](#) se aceptó el ajuste del modelo para valores de  $CFI \geq .95$ ,  $TLI \geq .96$ ,  $RMSR \leq .08$  y  $RMSEA \leq .08$ .

**Estimación y Ajuste del MCP.** El MCP (Modelo de Crédito Parcial) define la probabilidad que tiene una persona con un nivel de rasgo  $\theta$  de escoger la categoría  $h$  ( $h = 0, \dots, m$ ) en un ítem  $i$  con  $m + 1$  categorías. Para ello, se define el parámetro de umbral  $\beta_{ih}$ , que describe la cantidad de  $\theta$  que demanda el ítem para que el evaluado tenga más chances de elegir la categoría  $h$  en lugar de la categoría anterior ( $h - 1$ ). Si bien el MCP no condiciona la estimación de los  $\beta_{ih}$  de un mismo ítem para que presenten valores ordenados, es esperable que estos parámetros registren una secuencia creciente que refleje el incremento de las escalas Likert ([Embretson & Reise, 2000](#)).

Formalmente, la probabilidad de elegir  $h$  en el ítem  $i$  para un individuo con nivel de rasgo  $\theta$  sigue la función:

$$P_i(h|\theta) = \frac{e^{\sum_{k=0}^h \theta - \beta_{ik}}}{\sum_{j=0}^m e^{\sum_{n=0}^j \theta - \beta_{in}}}$$

para  $h = 0, \dots, m$ ;

Donde se define:  $\sum_{n=0}^0 (\theta - \beta_{in}) = 0$

La aplicación del MCP se realizó con el programa Winsteps ([Linacre, 2006](#)). Mediante el método de máxima verosimilitud conjunta se estimaron 32 parámetros de umbral para el formato de cinco opciones, y 40 parámetros para el for-

mato de seis opciones. Así también se estimaron dos valores de  $\theta$  para cada sujeto, extraídos sobre la base de sus patrones de respuestas a la prueba con los distintos formatos Likert. El ajuste global del modelo se estudió a partir de los estadísticos ajuste próximo (*Infit*) y lejano (*Outfit*), los cuales se interpretan como medias cuadráticas de los residuales no estandarizados (MNSQ). Estos indicadores adoptan un valor de 1 para un ajuste perfecto, y se contempla una zona de ajuste aceptable entre 0.5 y 1.5 tanto para el *Infit* como para el *Outfit* de los sujetos y de los ítems ([Linacre, 2012](#)).

**Asociación con otras variables.** Se analizó la asociación de diferentes criterios externos con los  $\theta$  estimados para cada individuo a partir de sus patrones de respuestas a los ítems de cinco categorías. Para criterios externos medidos cuantitativamente (edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática) se utilizaron coeficientes  $r$  de Pearson. En los casos en que el criterio externo era dicotómico (varones vs. mujeres y cursantes vs. recursantes de Matemática) se calcularon los coeficientes  $d$  de Cohen (y sus respectivos intervalos de confianza del 95%) para analizar el tamaño del efecto de la diferencia de medias. Idénticos análisis se usaron para estudiar la relación de los mismos criterios externos con los  $\theta$  estimados a partir las respuestas a los ítems con seis categorías. Posteriormente se procedió a contrastar los resultados obtenidos para cada criterio externo a fin de detectar diferencias estadísticamente significativas en la forma en que se relacionan con ambas estimaciones de  $\theta$ . Se aplicaron pruebas de diferencia de  $r$  de Pearson para muestras relacionadas ([Steiger, 1980](#)) con el fin de comparar los coeficientes de correlación obtenidos en el estudio de la asociación de cada criterio cuantitativo con los distintos formatos Likert. Los coeficientes  $d$  de Cohen fueron comparados considerando sus intervalos



**Tabla 1.**  
Supuesto de unidimensionalidad y análisis de consistencia interna.

	Escala Likert	
	5 categorías	6 categorías
Análisis Factorial Confirmatorio		
CFI	.978	.975
TLI	.970	.965
RMSEA [IC 99%]	.075 [.051 - .099]	.072 [.048 - .096]
RMSR	.044	.041
Consistencia interna		
Alfa de Cronbach [IC 99%]	.93 [.92, .94]	.92 [.91, .93]
<i>Greatest Lower Bound</i>	.95	.94

**Nota.** CFI = Índice de ajuste comparativo, TLI = Índice de Tucker-Lewis, RMSEA = Error medio cuadrático de aproximación, RMSR = Raíz del promedio de los residuos al cuadrado.

de confianza (IC). Siguiendo a [Maydeu-Olivares et al. \(2009\)](#), en este trabajo se asume que no hay diferencia significativa si se registra una superposición de los IC de las  $d$  de Cohen.

**Medidas de precisión.** Se estudió la consistencia interna del conjunto de ítems a partir de los coeficientes alfa de Cronbach y *Greatest Lower Bound* (GLB). En el marco de la TRI se examinaron las funciones de información de los tests para cada formato de respuesta, y la función de eficiencia relativa (FER). La eficiencia relativa se define como el *ratio* entre las funciones de información de cada formato, por lo que permite estimar el aporte de información que realiza un formato Likert por sobre otro.

## Resultados

### *Comparación del supuesto de unidimensionalidad*

Los índices de ajuste comparativo mostraron un resultado adecuado para la verificación del modelo unidimensional usando ambos formatos

de respuesta (TLI y CFI > .95). Con respecto a los índices de ajuste absoluto, los valores de RMSR fueron inferiores al criterio de .08, en tanto que los RMSEA reflejaron un cierto desajuste. Estos resultados permiten, en términos generales, aceptar razonablemente el cumplimiento del supuesto de unidimensionalidad bajo las condiciones que establecen ambas escalas Likert (Tabla 1).

### *Comparación de la estimación y el ajuste del MCP*

La estimación de los parámetros para las escalas Likert alcanzó el criterio de convergencia empleando una cantidad razonable de iteraciones. En la Tabla 2 se muestran los índices obtenidos para la evaluación global del ajuste del MCP a los datos de los dos formatos con cinco y seis categorías de respuesta. Para ambas escalas Likert, las medias de *Infit* y *Outfit* tanto para los sujetos como para los ítems resultaron próximas a 1, lo que refleja un ajuste aceptable.

**Tabla 2**  
Evaluación del ajuste global al Modelo de Crédito Parcial.

Escala Likert	Ajuste global de sujetos				Ajuste global de ítems			
	$\theta$	SE	Infit MNSQ	Outfit MNSQ	$\beta_i$	SE	Infit MNSQ	Outfit MNSQ
5 opciones								
Media	.21	.53	1.00	1.00	.00	.05	1.00	1.00
DE	1.80	.11	.42	.43	.47	.00	.08	0.10
6 opciones								
Media	.23	.43	1.03	1.02	.00	.04	1.01	1.02
DE	1.49	.12	.53	.53	.35	.00	.06	.07

**Nota.**  $\theta$  = Nivel de rasgo; SE = Error de estimación;  $\beta_i$  = promedio de los  $\beta_m$  de un ítem; MNSQ = Media cuadrática de los residuales no estandarizados del ajuste interno (*Infit*) y ajuste externo (*Outfit*).

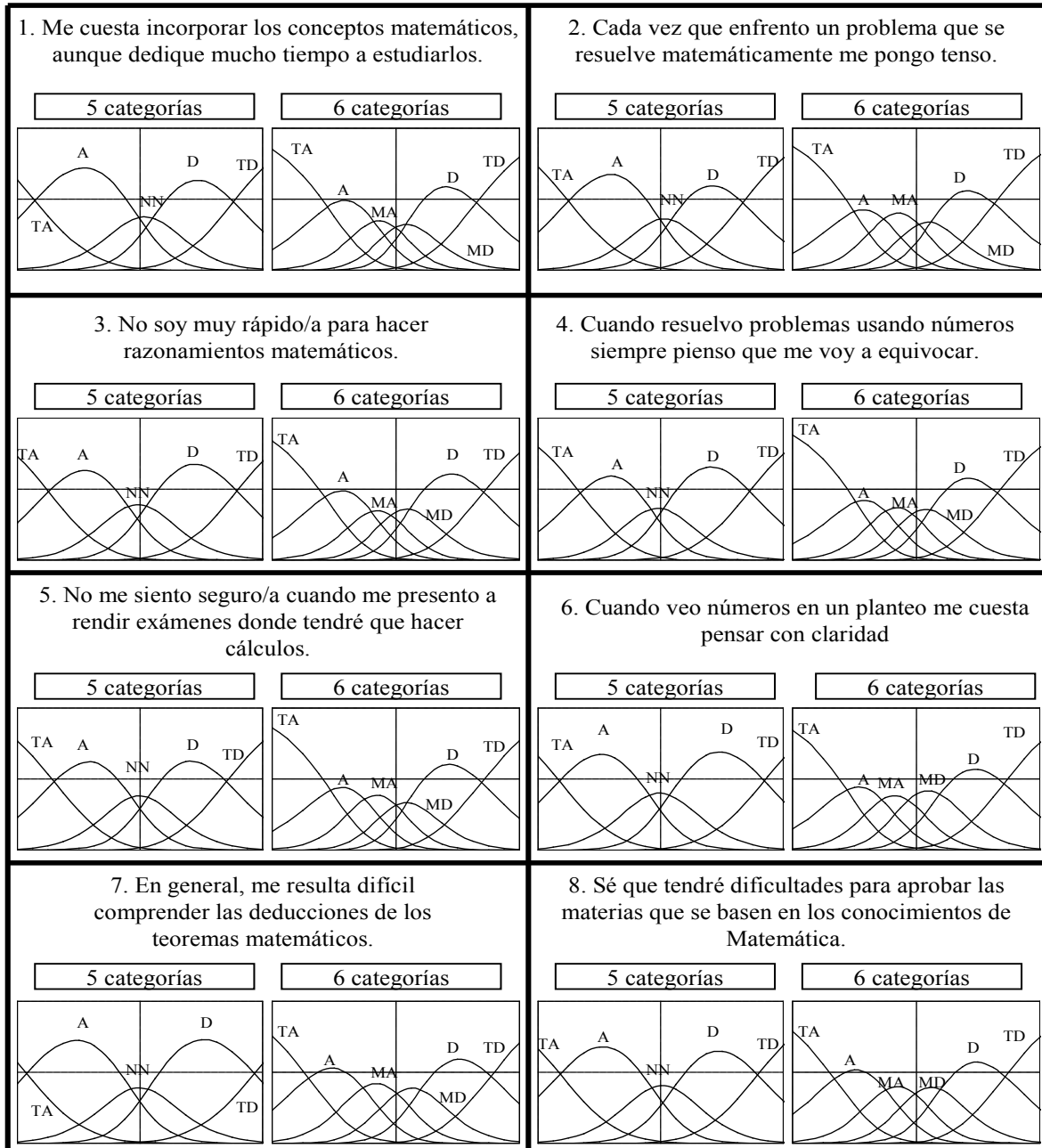
Se encontró una alta correlación entre los  $\theta$  estimados para cada formato, siendo  $r = .91$ ,  $p < .001$  (IC al 99%: .89, .92). Este resultado parece reflejar que el número de opciones podría incidir en la medición del constructo, aunque no sustancialmente. No se encontraron inversiones en los parámetros de umbral  $\beta_{ih}$  de los ítems estimados bajo las dos condiciones de formato. Todas las opciones de las escalas Likert permitieron discriminar algún rango específico del rasgo latente con independencia de la inclusión en la escala de una categoría central (Figura 1). Esto implica que los parámetros de cada ítem presentaron valores crecientes acordes con los niveles de acuerdo reflejados en los anclajes lingüísticos de las escalas Likert.

La categoría central en la escala Likert de cinco opciones mostró una tendencia a ser la más probable para los individuos con niveles de  $\theta$  intermedios. No obstante, el rango de valores del continuo del rasgo para los que la categoría intermedia mostró más probabilidad que el resto de las categorías de ser elegida resultó bastante acotado. Esto implica que incluso aquellos sujetos ubicados en niveles muy próximos a la media del rasgo también tienen altas probabilidades de optar por

las categorías adyacentes a la central. Como consecuencia, se observó en todos los ítems una clara polarización entre las opiniones de los estudiantes que estaban de acuerdo (categorías *De acuerdo* y *Totalmente de acuerdo*) y las de aquellos que estaban en desacuerdo (categorías *En desacuerdo* y *Totalmente en desacuerdo*). Al comparar estos resultados con las curvas características de los ítems que fueron respondidos con seis opciones se puede apreciar que no se modificó en gran medida la tendencia a elegir la categoría *En desacuerdo* y *Totalmente en desacuerdo*. En cambio, la propensión a estar *De acuerdo* en el formato de cinco categorías parece distribuirse de manera relativamente más equilibrada en las categorías *De acuerdo*, *Más bien de acuerdo* y *Más bien en desacuerdo* en el formato Likert de seis puntos.

#### *Comparación de la asociación con criterios externos*

Las asociaciones de los  $\theta$  con las variables externas resultaron acorde a lo esperado a nivel teórico. Como muestra la Tabla 3, los valores  $r$  de Pearson obtenidos al estudiar la correlación de la

**Figura 1**

Curvas características de los ítems respondidos con cinco y seis categorías.

confianza con otras variables (edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática) no mostraron diferencias estadísticamente significativas entre el formato de respuesta con categoría intermedia y el que no la incluía. Tampoco se encontraron diferencias al

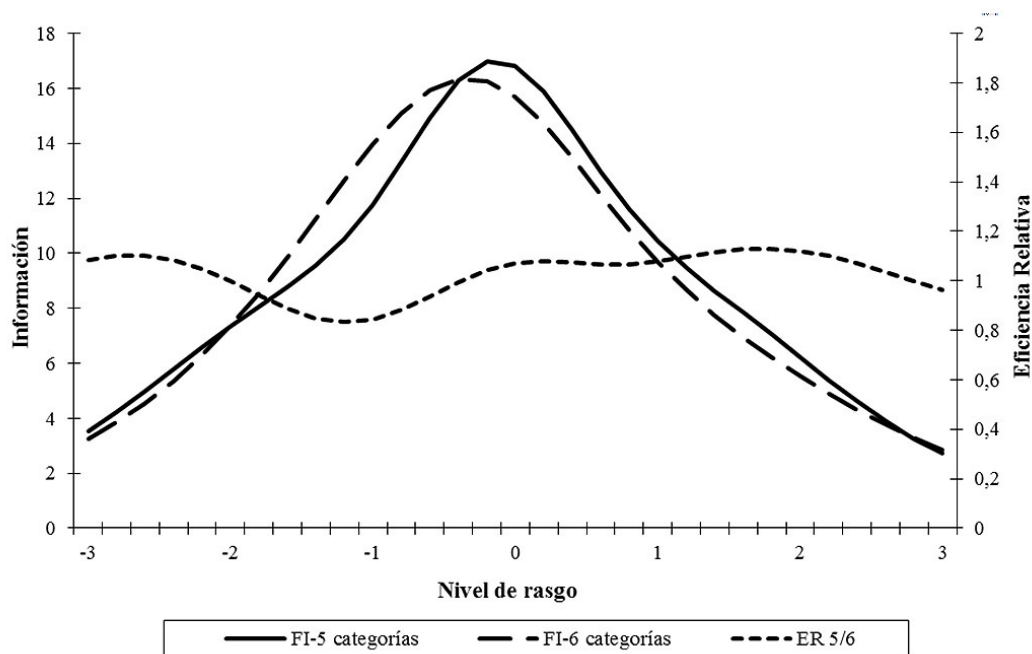
contrastar grupos de datos en base a criterios dicotómicos (varones vs. mujeres, cursantes vs. re-cursantes de Matemática). En efecto, se encontró una superposición entre los IC de los tamaños del efecto  $d$  de Cohen de los dos formatos.



**Tabla 3**

Relación de utilidad con criterios externos.

Criterio externo	Estadístico	Escala Likert		Significación
		5 categorías	6 categorías	
Sexo	$d$ [95% IC]	.37 [.24, .50]	.32 [.20, .41]	---
Edad	$r$	-.023	-.001	$t_{(936)} = -1.57$ , NS.
Recurso Matemática	$d$ [95% IC]	.52 [.39 - .64]	.51 [.39 - .60]	---
Nota en Matemática	$r$	.266***	.278***	$t_{(936)} = -0.89$ , NS.
Afecto hacia la Matemática	$r$	.675***	.690***	$t_{(936)} = -1.49$ , NS.
Utilidad de la Matemática	$r$	.310***	.326***	$t_{(936)} = -1.21$ , NS.

Nota. \*\*\* $p < .001$ **Figura 2**

Funciones de información para los formatos de cinco y seis categorías, y función de eficiencia relativa.

### Comparación de medidas de precisión

Los índices de consistencia interna alfa de Cronbach y GLB registraron valores altamente satisfactorios y similares para ambos formatos de escala Likert (Tabla 1). Las funciones de información de los tests muestran que el instrumento aporta mayor cantidad de información en los ni-

veles medio y medio-bajo en ambos modelos, de cinco o seis categorías. Este resultado se vio reflejado en la FER, que se mantuvo relativamente constante y próxima a 1 a lo largo de todo el espectro del rasgo latente, lo que indica que la precisión de medición es similar en ambas versiones del test (Figura 2).

## Discusión

La especificación del formato de respuesta de los ítems pertenecientes a tests de comportamiento típico supone un conjunto de decisiones que debe tomar quien construye un instrumento. A pesar de esto, es poco frecuente que los autores justifiquen por qué utilizan una cantidad particular de opciones o por qué escogen determinados cuantificadores lingüísticos para acompañar las categorías. Más infrecuente es encontrar que los autores recurran a estudios empíricos para analizar si estas decisiones adoptadas en el diseño de la prueba pueden perjudicar la medición. El objetivo de este trabajo fue estudiar si la inclusión o no de una categoría central en la escala Likert de la Escala de Confianza para la Matemática impacta sobre las propiedades psicométricas del instrumento.

Las diferencias encontradas al analizar las evidencias de validez basadas en la estructura interna no resultaron considerables. Los índices de bondad de ajuste del AFC adoptaron valores óptimos y similares para las mediciones de *confianza* obtenidas a partir de ítems con formato tanto de cinco como de seis categorías. En la misma línea, tampoco se observó una discrepancia en la evaluación del ajuste al MCP. Algunos autores han encontrado que una disminución del número de opciones parece asociarse a una mejora en el ajuste de los modelos de la TRI (Hernández, Muñiz, & García-Cueto, 2000; Lee & Paek, 2014; Maydeu-Olivares et al., 2009). Sin embargo este efecto no se ha percibido en el presente estudio, posiblemente debido a que se torna más notable en la medida en que el ítem presenta menos cantidad de categorías (p. ej. tres o cuatro opciones).

El cambio en la escala Likert tampoco afectó la relación de la *confianza* con otras variables consideradas como criterios externos. Los tamaños de los efectos de las diferencias de medias de

*confianza* entre varones y mujeres fueron semejantes cuando el constructo se midió usando los ítems con cinco o con seis categorías. Este mismo resultado se corroboró en los tamaños de los efectos en la comparación de los grupos de estudiantes que aprobaron Matemática la primera vez que la cursaron y los que debieron recursarla. Por otra parte, las asociaciones de la *confianza* con las variables edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática también presentaron resultados similares al cotejar los coeficientes de correlación obtenidos para cada formato de respuesta.

Con respecto a la confiabilidad, la inclusión de la categoría central no produjo un efecto significativo en los índices de consistencia interna calculados en el marco de la teoría clásica. Desde la perspectiva de la TRI, la FER mostró que usando ambos formatos de respuesta se alcanza una precisión similar para todos los niveles del rasgo.

En conclusión, los resultados obtenidos permiten afirmar que el uso de formatos de respuesta con cinco o seis categorías no afecta sustancialmente las evidencias de validez ni confiabilidad de la Escala de Confianza para la Matemática.

El análisis de los ítems de la Escala de Confianza mediante el MCP mostró que la categoría central es eficaz para detectar sujetos que presentan niveles medios del rasgo. En efecto, para todos los ítems se observó un rango de valores dentro del espectro del rasgo en el que las curvas características de la opción *Ni de acuerdo ni en desacuerdo* fueron más probables que las curvas de las demás categorías del ítem. En un estudio previo, empleando el mismo anclaje lingüístico, Rojas y Fernández (2000) encontraron un comportamiento de la curva de la categoría similar al aquí observado en un grupo de universitarios que interpretaban esta opción como una categoría que expresa una posición intermedia. En cambio, un grupo que interpretaba la categoría central en tér-

minos de duda o indecisión mostró parámetros de umbral desordenados y, consecuentemente, curvas de categorías que no resultaban las más probables en ningún intervalo de la escala. La homogeneidad en los resultados obtenidos con el grupo de estudio en los distintos ítems apoya la hipótesis de una tendencia de los sujetos a atribuir un significado acorde a lo esperado para la categoría central de la escala. No obstante, nos parece conveniente complementar este resultado con estudios cualitativos que permitan corroborar los patrones de interpretación de las opciones. Al igual que en los estudios realizados por [Kulas y Stachowski \(2013\)](#), aquí también se podría solicitar a un grupo de los sujetos que informen *en voz alta* el proceso que lleva a la elección de una respuesta al ítem.

Al observar las curvas de los ítems respondidos con la escala Likert de cinco opciones podría aceptarse que la categoría central resulta eficaz en tanto que atrae a una parte no despreciable de sujetos. Sin embargo, para todos los ítems se verifica que la probabilidad de elección de la categoría central disminuye considerablemente al ser comparada con las probabilidades de optar por las categorías adyacentes. Esto implica que, aún teniendo la posibilidad de escoger una respuesta neutra, la mayoría de los alumnos tienden a expresar una posición a favor o en contra de la opinión planteada en el ítem. Posiblemente este resultado se encuentre relacionado con el constructo medido. La confianza del estudiante sobre su destreza matemática, en tanto creencia, se forma sobre la base de su trayectoria académica ([Nuñez et al., 2005](#)). En consecuencia, es esperable que al alcanzar un nivel universitario, el individuo se afirme en sus experiencias previas a fin de percatarse de las expectativas de logro a las que se podría enfrentar, y logre discernir con mayor claridad, al menos en términos dicotómicos, si puede alcanzarlas o no.

Ahora bien, la disminución en la probabili-

dad de elección de la categoría central indica que podría resultar adecuado ensayar un formato de cuatro opciones. En esta línea, en futuros estudios buscaremos analizar si una escala Likert más parsimoniosa, de cuatro categorías, afecta significativamente el grado de ajuste del modelo o la precisión de la medida. Así también, examinaremos si los resultados obtenidos en este trabajo son comparables con los de grupos de universitarios de otras carreras y para la medición de otros constructos.

En cuanto a las limitaciones del presente estudio, conviene señalar que las conclusiones obtenidas sobre el funcionamiento de la categoría central se circunscriben estrictamente a los ítems de la Escala de Confianza para la Matemática. Aún cuando las posibilidades de generalización de estos resultados sean reducidas, los mismos podrían considerarse orientadores respecto a otras escalas que midan un constructo afin o que indaguen en una población con características similares. El diseño metodológico con medidas repetidas implementado en este estudio permite analizar los efectos de variaciones intra-individuales en cuanto a la respuesta de un mismo ítem bajo dos formatos de respuesta diferentes: de cinco y de seis opciones. Esto resulta una ventaja frente a abordajes con datos simulados, los cuales examinan el efecto de la categoría intermedia sobre la validez y la confiabilidad del instrumento, pero sin considerar el proceso psicológico subyacente involucrado en la respuesta a los ítems en un test específico.

## Referencias

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Abal, F., Auné, S., & Attorresi, H. (2014). Comparación del modelo de respuesta graduada y la teoría clásica

- sica de tests en una Escala de Confianza para la Matemática. *Summa Psicológica UST*, 11(2), 101-113. Recuperado de <http://summapsicologica.cl/index.php/summa>
- Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2015). Modelización de una prueba de afecto hacia la matemática con la teoría de respuesta al ítem. *Revista de Psicología UCA*, 11(21), 23-34. Recuperado de <http://bibliotecadigital.uca.edu.ar>
- Abal, F. J. P., Galibert, M. S., Aguerri, M. E., & Attorresi, H. F. (2014). Comparación de los modelos respuesta graduada y crédito parcial aplicados a una escala de utilidad de la matemática. *Revista Argentina de Ciencias del Comportamiento*, 6(3), 6-16. Recuperado de <https://revistas.unc.edu.ar/index.php/racc>
- Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey: Measuring third through sixth graders' attitudes towards mathematics. *Measurement and Evaluation in Counseling and Development*, 44(4), 225-247. doi: 10.1177/0748175611418522
- Andrich, D., de Jong, J. H. A. L., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. En J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 59-72). Münster, Germany: Waxmann Verlag.
- Auzmendi-Escribano, E. (1992). *Las actitudes hacia la matemática-estadística en las enseñanzas medias y universitarias: Características y medición*. Bilbao, España: Mensajero.
- Bazán, J. L., & Sotero, H. (1998). Una aplicación al estudio de actitudes hacia la matemática en la UNALM. *Anales Científicos UNALM*, 36, 60-72.
- Bermejo, V. (1996). Enseñar a comprender las matemáticas. En J. Beltrán-Llera & C. Genovard-Roselló (Eds.), *Psicología de la Instrucción I* (pp. 256-279). Madrid: Síntesis.
- Bisquerra, R., & Pérez-Escoda, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *Revista d'Innovació i Recerca en Educació*, 8(2), 129-147. doi: 10.1344/reire2015.8.2828
- Blanco-Blanco, A. (2004). Enseñar y aprender estadística en las titulaciones universitarias de ciencias sociales: Apuntes sobre el problema desde una perspectiva pedagógica. En J. C. Torre-Puente & E. Gil-Coria (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje* (pp. 143-190). Madrid: Universidad Pontificia Comillas.
- Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7(5), 324-326. doi: 10.2307/748467
- Gil-Ignacio, N., Guerrero-Barona, E., & Blanco-Nieto, L. (2006). El dominio afectivo en el aprendizaje de las matemáticas. *Electronic Journal of Research in Educational Psychology*, 4(8), 47-72. Recuperado de <http://investigacion-psicopedagogica.org/revista/new/index.php>
- Gómez-Chacón, I. M. (2005). *Matemática emocional. Los afectos en el aprendizaje matemático*. Madrid, España: Narcea.
- Gómez-Chacón, I. M., Op't Eynde, P., & De Corte, E. (2006). Creencias de los estudiantes de matemáticas. La influencia del contexto de clase. *Enseñanza de las Ciencias*, 24(3), 309-324. Recuperado de <http://ddd.uab.cat/record/18?ln=carecord/14?ln=ca>
- González-Romá, V., & Espejo, B. (2003). Testing the middle response categories "Not sure", "In between" and "?" in polytomous items. *Psicothema*, 15(2), 278-284. Recuperado de <http://www.psicothema.es/>
- Hernández, A., Drasgow, F., & González-Romá, V. (2004). Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology*, 89(4), 687-699. doi: 10.1037/0021-9010.89.4.687
- Hernández, A., Espejo, B., & González-Romá, V. (2006).

- The functioning of central categories Middle Level and Sometimes in graded response scales: Does the label matter? *Psicothema*, 18(2), 300-306. Recuperado de <http://www.psicothema.es/>
- Hernández-Baeza, A., Espejo-Tort, B., González-Romá, V., & Gómez-Benito, J. (2001). Escalas de respuesta tipo Likert: ¿Es relevante la alternativa “indiferente”? *Metodología de Encuestas*, 3(2), 135-150. Recuperado de <http://casus.usal.es/pkp/index.php/MdE>
- Hernández, A., Muñiz, J., & García-Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema*, 12(2), 288-291. Recuperado de <http://www.psicothema.es/>
- Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties*, 15(2), 237-264, doi: 10.1080/13689880500178849
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert-scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403. doi: 10.9734/BJAST/2015/14975
- Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43(3), 489-493. doi: 10.1016/j.jrp.2008.12.005
- Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing: Person and item contributors to middle category endorsement intent on Likert personality indicators. *Journal of Research in Personality*, 47(4), 254-262. doi: 10.1016/j.jrp.2013.01.014
- Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, 32(7), 663-673. doi: 10.1177/0734282914522200
- Linacre, J. M. (2006). *Winsteps® (Version 3.63.0)* [Software]. Beaverton, Oregon: Winsteps.com.
- Linacre, J. M. (2012). *Winsteps® Rasch measurement computer program User's Guide*. Beaverton, Oregon: Winsteps.com.
- Masters, G. N. (2016). Partial credit model. En W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume 1: Models* (pp. 109-126). Boca Raton, FL: CRC Press.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674. doi: 10.1177/001316447103100307.
- Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, 41(2), 295-308. doi: 10.3758/BRM.41.2.295
- McLeod, D. B., & McLeod, S. H. (2002). Synthesis - Beliefs and mathematics education: Implications for learning, teaching and research. En G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 115-126). Dordrecht: Kluwer Academic Publishers.
- Morales, P. M. (2006). *Medición de actitudes en psicología y educación* (3º ed.). Madrid: Universidad Pontificia Comillas.
- Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences*, 38(1), 61-69. doi: 10.1016/j.paid.2004.03.021
- Murray, A. L., Booth, T., & Molenaar, D. (2016). When middle really means “Top” or “Bottom”: An analysis of the 16PF5 using Bock's nominal response model. *Journal of Personality Assessment*, 98(3), 319-331. doi: 10.1080/00223891.2015.1095197
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide* (6º ed.). Los Angeles, CA: Muthén & Muthén.
- Narro-Ramírez, A. E. (1997). Investigación sobre la concepción de la matemática en las ciencias sociales

- en la UAM-Xochimilco. *Política y Cultura*, 9, 249-280. Recuperado de <http://www.redalyc.org/revista.oa?id=267>
- Nunes, C. H. S. S., Primi, R., Farias-Oliveira-Nunes, M., Muniz, M., Freitas-da Cunha, T., & Couto, G. (2008). Teoría de respuesta ao item para optimização de escalas tipo Likert –um exemplo de aplicação. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 25(1), 51-79. Recuperado de <http://www.redalyc.org/revista.oa?id=4596>
- Núñez, J. C., González-Pienda, J. A., Alvarez, L., González, P., González-Pumariega, S., Roces, C. ... Rodrigues-Feio, L. do S. (2005). Las actitudes hacia las matemáticas: perspectiva evolutiva. *Actas del VIII Congresso Galaico Português de Psicopedagogia*, 2389-2396. Recuperado de <http://www.educacion.udc.es/grupos/gipdae/index.php?pagina=VIIIcongreso>
- Palacios, A., Arias, V., & Arias, B. (2014). Attitudes towards mathematics: Construction and validation of a measurement instrument. *Revista de Psicodidáctica*, 19(1), 67-91. doi: 10.1387/RevPsicodidact.8961
- Preston, K., Reise, S., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS emotional distress item pools. *Educational and Psychological Measurement*, 71(3), 523-550. doi: 10.1177/0013164410382250
- Rojas-Tejada, A. J., & Fernández-Prados, J. S. (2000). Análisis de las alternativas de respuestas intermedias mediante el modelo de escalas de clasificación. *Metodología de Encuestas*, 2(2), 171-183. Recuperado de <http://casus.usal.es/pkp/index.php/MdE>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251. doi: 10.1037/0033-2909.87.2.245
- Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8(2), 16-21. Recuperado de <http://www.rapidintellect.com/AEQweb/cho253441.htm>
- Thissen, D., & Cai, L. (2016). Nominal categories models. En W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume 1: Models* (pp. 51-73). Boca Raton, FL: CRC Press.
- Tsang, K. K. (2012). The use of midpoint on Likert Scale: The implications for educational research. *Hong Kong Teachers' Centre Journal*, 11, 121-130. Recuperado de <http://www.edb.org.hk/hkct/content.aspx?id=20060614161412&lang=e>