

Artículo Original

Abal, Facundo Juan Pablo *^a, Galibert, María Silvia^a, Aguerri, María Ester^a, y Attorresi, Horacio Félix^a.

Resumen

Se aplicaron los Modelos Respuesta Graduada (MRG) y Crédito Parcial (MCP) de la Teoría de Respuesta al Ítem (TRI) al análisis de ítems de una escala que mide las creencias sobre la utilidad de la Matemática de estudiantes de Psicología. La prueba consta de 8 ítems con formato de respuesta Likert de 6 opciones. Participaron 1875 estudiantes de Psicología de la Universidad de Buenos Aires. Todos los análisis basados sobre la TRI se realizaron con el programa MULTILOG. La estimación de los parámetros de los modelos se efectuó por Máxima Verosimilitud Marginal. Previamente se verificó la condición de unidimensionalidad requerida por los modelos. El MRG presentó mejores indicadores de ajuste que el MCP y una Función de Información del Test más elevada. Los resultados aportan evidencias de validez basadas en la estructura interna del instrumento.

Palabras claves:

Utilidad de la Matemática; Creencias sobre Matemática; Modelo de Respuesta Graduada; Modelo de Crédito Parcial; Teoría de Respuesta al ítem.

Recibido el 8 de Mayo de 2014; Recibido la revisión el 19 de Septiembre de 2014; Aceptado el 22 de Octubre de 2014.

Abstract

A comparison of the Graded Response and Partial Credit models with a scale of Usefulness of Mathematics: The Graded Response Model (GRM) and the Partial Credit Model (PCM) of Item Response Theory (IRT) were applied in the item analysis of a scale that evaluates psychology students' beliefs about the usefulness of mathematics. The scale comprises 8 items in polytomous response format (6-point Likert-type). The sample was made up by 1875 students of the Psychology school of the University of Buenos Aires. All analyses based on IRT were performed by operating the MULTILOG software. The parameters estimation was carried out through Marginal Maximum Likelihood procedures. Previously, the unidimensionality assumption required by the models was confirmed. The GRM showed better fit indexes than PCM and higher Test Information Function. The findings provide validity evidences based on the internal structure of scale.

Key Words:

Usefulness of mathematics; Beliefs about mathematics; Graded Response Model; Partial Credit Model; Item Response Theory.

Tabla de Contenido

Introducción	6
Método	9
Participantes	9
Instrumento	9
Procedimiento	9
Análisis de	10
Datos	
Resultados	10
Discusión	12
Agradecimientos	14
Referencias	14

1. Introducción

Las creencias de los estudiantes universitarios sobre la utilidad que tiene la Matemática o la Estadística en su desarrollo académico-profesional constituyen uno de los aspectos más investigados en el campo de la dimensión afectiva que interviene en sus aprendizajes (Bermejo, 1996; Gómez Chacón, 2005; Hernández, 2011; McLeod & McLeod, 2002). Los alumnos se predisponen mejor a cursar estas materias y adoptan una postura más receptiva en la medida

que alcanzan una idea más clara de la aplicabilidad de las nociones que son objeto de estudio (Gil, García, Rodríguez, Guijarro & López, 1995; Ponsoda, 1990).

Los orígenes del estudio de estas creencias remiten al componente cognitivo de las concepciones tradicionales sobre actitud hacia la Matemática. La mayoría de las definiciones de actitud convergen en caracterizarla como un juicio evaluativo que se realiza sobre un objeto, en este caso, la Matemática y que

^a Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires, Argentina

*Enviar correspondencia a: Abal, F. J. P. mail: fabal@psi.uba.ar

contempla los dominios cognitivo, afectivo y comportamental (Gómez Chacón, 2005). Específicamente, el componente cognitivo se refiere al conjunto de creencias y opiniones que la persona tiene respecto de la Matemática (Martínez, 2008). Son numerosos los modelos teóricos sobre la actitud hacia la Matemática que incluyen la evaluación de la información con la que cuenta el estudiante acerca de los beneficios y productividad de adquirir conocimiento matemático para su propia vida académica y/o profesional. Esta subdimensión de la actitud ha recibido numerosas denominaciones: Utilidad (Auzmendi, 1992; Fennema & Sherman, 1976; Palacios, Arias & Arias, 2014), Aplicabilidad (Bazán & Sotero, 1998; Rodríguez Feijóo, 2011) y Valor (Tapia & Marsh, 2004).

En las últimas décadas las creencias han pasado a recibir mayor atención por parte de los investigadores como uno de los descriptores básicos para entender la dimensión afectiva de la Matemática (Goldin, Roesken & Törner, 2009; Gómez Chacón, 2005, 2010). Las creencias sobre la Matemática son una parte del conocimiento subjetivo que el individuo ha construido sobre esta materia a partir de las experiencias de enseñanza y aprendizaje que ha atravesado. Se trata de un conjunto arraigado de ideas relacionadas entre sí, difíciles de modificar y con una gran estabilidad en el tiempo (Gil, Blanco & Guerrero, 2006).

En el marco de los estudios sobre las creencias preponderantes del aprendizaje de la Matemática, se han identificado múltiples objetos de estas creencias: acerca de la naturaleza de la Matemática, de uno mismo como aprendiz, de las prácticas de enseñanza y del contexto social en el que acontece la educación matemática (McLeod & McLeod, 2002). Particularmente, Bermejo (1996) destacó que las creencias acerca de la naturaleza de la Matemática se vinculan principalmente con la percepción de importancia y utilidad que el estudiante le atribuye a esta disciplina. Se ha encontrado que la toma de conciencia acerca de la de utilidad de la Matemática correlaciona positivamente con el rendimiento académico en tanto que promueve conductas de interés, esfuerzo, perseverancia y disposición en su aprendizaje (Álvarez & Ruiz, 2010; Gil et al., 2006).

Ahora bien, resulta notorio que existen diferencias entre las creencias de utilidad de la

Matemática en función de la rama del conocimiento que se estudie. Mientras que en las carreras de ciencias físico-químico-matemáticas o tecnológicas estas creencias aparecen como las más establecidas (Álvarez & Ruiz, 2010; Martínez & Gaitán, 2013), se acepta como una realidad universal que una parte considerable de los alumnos de carreras humanísticas presentan una actitud desfavorable frente al aprendizaje de materias de corte cuantitativo (Coolican, 1994; Rodríguez Feijóo, 2011). En carreras como Psicología, Sociología y Ciencias de la Educación es indispensable la implementación de estrategias didácticas que modifiquen el rechazo o la indiferencia de los estudiantes para posibilitar un aprendizaje significativo (Cortada, Rodríguez Feijóo & Kohan Cortada, 2008).

Las características particulares de esta población de estudiantes universitarios obliga a pensar en la necesidad de instrumentos de medición más precisos que permitan detectar con más detalle las diferencias individuales. Sobre la base de las referencias revisadas se estableció una delimitación conceptual del constructo Utilidad de la Matemática adaptado para estudiantes de carreras humanístico-sociales. Se definió operacionalmente como un conjunto de creencias con respecto a la relación entre la disciplina científica estudiada y la Matemática y de la aplicabilidad de las nociones matemáticas durante la formación académica y en el futuro desempeño profesional (Abal, Lozzia, Aguerri, Galibert & Attorresi, 2010b).

Vale aclarar que las creencias sobre la Matemática no son las únicas que intervienen en el proceso de enseñanza-aprendizaje. Bermejo (1996) consideró además la importancia de las creencias vinculadas a la capacidad percibida por el alumno para enfrentar exitosamente situaciones en donde se aplique matemática. Éstas tienen una fuerte carga afectiva y se ha demostrado que son mejores predictores del rendimiento académico (Gil et al., 2006; Gómez Chacón, 2005).

No obstante, en el presente estudio se ha priorizado el desarrollo de la escala de Utilidad porque investigaciones previas con esta variable evidenciaron que, aunque la correlación con el rendimiento puede ser modesta, resultan más fáciles de modificar contemplando las condicionantes de tiempo y masificación que presentan los cursos

universitarios. En cambio, las creencias vinculadas a la confianza en sí mismo para hacer frente a la Matemática tiene una mayor estabilidad y resistencia al cambio (Abal, 2013; Abal et al., 2010b).

La presente investigación tiene como objetivo buscar evidencias de validez basadas sobre la estructura interna de una escala construida por los autores que mide la Utilidad de la Matemática para población de estudiantes de Psicología (Abal, 2013). Para tal fin se plantea una comparación de las aplicaciones de dos modelos politómicos de la Teoría de Respuesta al Ítem (TRI) a los ítems del instrumento: Modelo de Respuesta Graduada (MRG) de Samejima (1969) y Modelo Crédito Parcial (MCP) de Master y Wrigth (1997). Cada modelo es una formulación matemática que plantea explicaciones plausibles sobre cómo se vincula la puntuación observada de un evaluado en un ítem con la cantidad de rasgo que posee. Se trata de una construcción hipotética sostenida por una serie de supuestos que, si se cumplen, garantizan la calidad de la medida (Muñiz, Fidalgo, García Cueto, Martínez & Moreno, 2005).

El MRG permite calcular la probabilidad que tiene una persona con un nivel de rasgo θ de elegir la categoría j ($j = 0, \dots, m$) en el ítem i de $m+1$ opciones a partir de dos pasos. En el primero se calculan las probabilidades de optar por la categoría j o una superior según:

Para $j = 0$ se define $P_{i0}^* = 1$

Para $j = 1, \dots, m$ se define $P_{ij}^*(\theta) = \frac{1}{1 + e^{-1.7a_i(\theta - b_{ij})}}$

Para $j = m+1$ se define $P_{i(m+1)}^* = 0$

Donde a_i es el parámetro de discriminación de cada ítem y b_{ij} , definido solamente para $j = 1, \dots, m$, es una serie de parámetros de umbral que separan las categorías adyacentes de la escala graduada. El siguiente paso consiste en obtener las denominadas curvas características de las categorías de los ítems a partir de una resta de las probabilidades acumuladas a derecha:

$$P_{ij}(\theta) = P_{ij}^*(\theta) - P_{i(j+1)}^*(\theta) \text{ para } j = 0, \dots, m$$

El MCP deriva su nombre de su originaria aplicación a la evaluación del rendimiento en el ámbito educativo. Este modelo permite estudiar el comportamiento de ítems que ponderan el conocimiento parcial del examinado; ya sea porque

aceptan respuestas parcialmente correctas y/o porque reciben calificaciones graduadas en varios niveles de calidad. Aunque fue creado para la medición del rendimiento y habilidades, Masters y Wright (1997) aseguraron que, por sus características, el MCP puede también aplicarse en el campo de la evaluación de las actitudes y la personalidad si los ítems son de puntuación ordinal.

El MCP permite calcular la probabilidad que tiene un individuo de elegir la categoría h ($h = 0, \dots, m$) en el ítem i de $m+1$ opciones dado su nivel de rasgo θ . Se trata de una generalización del clásico Modelo de Rasch (Rasch, 1960) y contempla la aplicación de este modelo dicotómico a partir de una segmentación adyacente del dato politómico. Esto es, se focaliza en la cantidad de rasgo que demanda el ítem para que el evaluado tenga más chances de elegir la categoría h en lugar de la inmediatamente anterior ($h - 1$).

La transición o paso de una categoría a la otra dependen de si el individuo cuenta con un nivel de θ suficiente como para superar un parámetro de umbral β_{ih} correspondiente a cada paso. La denominación *paso* es heredada de la aplicación del MCP en tests de comportamiento máximo. En ítems de tests de comportamiento típico, este paso como tal, debe suponerse debido a que no se obtiene un registro tangible de la transición como podría ser un ítem de habilidad parcialmente resuelto. Cada β_{ih} se define como la localización del paso en la escala del rasgo. Se trata de un valor de umbral en el que el evaluado tiene la misma probabilidad de responder en la categoría h o en $h - 1$. Lógicamente, este parámetro está definido solamente para $h = 1, \dots, m$ porque no existe una categoría anterior a $h = 0$ y, por ende, tampoco existe el umbral β_{i0} . Esto significa que si un ítem tiene $m+1$ categorías, el individuo debería dar m pasos para elegir la opción más elevada de la escala Likert.

Formalmente, un evaluado de rasgo θ tiene una probabilidad de elegir h en el ítem i según la ecuación:

$$P_i(h|\theta) = \frac{e^{\sum_{k=0}^h \theta - \beta_{ik}}}{\sum_{j=0}^m e^{\sum_{h=0}^j \theta - \beta_{ih}}} \text{ para } h = 0, \dots, m$$

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

Conviene resaltar que, a diferencia del MRG, el MCP supone la homogeneidad de la capacidad discriminativa de los ítems. Básicamente, el modelo define que la probabilidad que tiene una persona de escoger h , se corresponde con la diferencia entre su nivel de rasgo θ y el parámetro de umbral β_h asociado a cada categoría, considerando todos los pasos del ítem en conjunto (Gempp et al., 2006). El modelo representa la probabilidad de responder a una categoría h como una exponencial correspondiente a esta categoría que se divide por la sumatoria de las exponenciales de todas las categorías (Asún & Zuñiga, 2008; Embretson & Reise, 2000; Martínez Arias, Hernández Lloreda & Hernández Lloreda, 2006).

La comparación de las aplicaciones de distintos modelos es una estrategia metodológica frecuente desde la perspectiva de la TRI (Asún & Zuñiga, 2008; Kang & Cohen, 2007; Reise & Waller, 2003). Esto permite determinar cuál es la estructura paramétrica más adecuada para la modelización de la variable estudiada. Al mismo tiempo, ofrece una valiosa comprensión del rasgo medido a partir de análisis exhaustivo de los parámetros de los ítems y de las medidas locales de precisión.

2. Método

2.1. Participantes

Se contó con la colaboración de una muestra incidental de 1875 alumnos que cursaban la materia Estadística de la carrera de Psicología de la Universidad de Buenos Aires. El 83% eran mujeres y el 17% eran varones. La edad promedio de los participantes fue de 22.7 años ($DE = 6.33$) y osciló entre 18 y 62 años.

2.2. Instrumento

Escala de Utilidad de la Matemática. La escala administrada para esta investigación ha sido sometida a un proceso de depuración realizado en etapas previas (Abal, 2013; Abal et al., 2010b). Se partió de un conjunto inicial de 25 ítems elaborados a partir de indicadores procedentes de diversos instrumentos elaborados por otros autores que miden constructos similares (Auzmendi, 1992; Bazán & Sotero, 1998) y se adicionaron indicadores originales surgidos de entrevistas grupales con estudiantes de

características afines a los de la población meta. Se puso énfasis en que el contenido del ítem haga una mención explícita de la aplicación de nociones matemáticas en la carrera de Psicología. Los ítems generados fueron analizados por jueces expertos de psicoestadística, psicometría y evaluación psicológica. Los jueces desestimaron un conjunto de ítems que apelaban a una visión histórica del vínculo entre la Psicología y la Matemática y que difícilmente podrían ser respondidos por estudiantes en el ciclo inicial de la carrera. Posteriormente se realizaron dos estudios piloto que mostraron la adecuación del vocabulario empleado y sirvieron para realizar una depuración de los ítems basada en indicadores de la teoría clásica.

La versión actual de la escala consta de ocho enunciados con respuesta tipo Likert de seis opciones: *Totalmente en desacuerdo*, *En desacuerdo*, *Más bien en desacuerdo*, *Más bien de acuerdo*, *De acuerdo* y *Totalmente de acuerdo* (Figura 1). Los estudiantes con elevados niveles en esta variable consideran que la Matemática es una disciplina que enriquece la formación del psicólogo y es indispensable para el futuro profesional. En cambio, los estudiantes con bajos puntajes opinan que la formación matemática de la carrera de Psicología debiera estar acotada al interés particular del grupo de alumnos que busque especializarse. Consideran que la Matemática tiene una estructura rígida que no es aplicable para analizar los fenómenos psicológicos. La consistencia interna de la prueba medida mediante el Alfa de Cronbach fue de .89. También se estudió la estabilidad temporal de las puntuaciones a partir de la técnica test-retest con intervalo de 15 días. Para este estudio se consideró una submuestra aleatoria de individuos ($n = 63$) que presentó un coeficiente $r = .83$, $p < .001$ entre los puntajes de ambas administraciones.

2.3. Procedimiento

Se adoptó un diseño acorde con los requisitos de un estudio instrumental de corte psicométrico (Montero & León, 2005). Los estudiantes que respondieron el cuestionario de forma anónima y voluntaria. La administración se realizó la tercera semana del curso de Estadística para asegurar que los estudiantes hubiesen tenido una aproximación al uso de algunas nociones de Matemática en el análisis de datos en Psicología.

2.4. Análisis de Datos

Se corroboró el supuesto de unidimensionalidad del espacio latente a partir de un Análisis de Componentes Principales. Para la aplicación de los MRG y MCP se recurrió al software MULTILOG (Thissen, 2003). La estimación de los parámetros para ambos modelos se realizó por Máxima Verosimilitud Marginal. Como indicador del ajuste de los datos al modelo, el programa brinda proporciones observadas y esperadas de elección para cada categoría de respuesta. Dado que esta información como única evidencia es limitada (Reuelta, Abad & Ponsoda, 2006), se adicionaron indicadores indirectos de ajuste recomendados por otros investigadores (Gray-Little, Williams & Hancock, 1997; Rubio, Aguado, Hontangas & Hernández, 2007): a) convergencia en la estimación en una cantidad de iteraciones razonables; b) estimación de parámetros con valores acordes con lo esperado, c) errores de estimación relativamente bajos y d) evidencia empírica de la invarianza de los parámetros. Aunque existen otros software que proponen indicadores más estrictos para estudiar el ajuste de los datos se consideró pertinente, como el objetivo central del trabajo es la comparación de las aplicaciones de ambos modelos, emplear el mismo programa para ambas aplicaciones y, de esta manera, homogeneizar los procedimientos de estimación y

evaluación de ajuste. De otra forma no podría concluirse respecto de cuál es la modelización más adecuada si la evaluación del ajuste se realiza sobre criterios más estrictos para un modelo y más débiles para otro. Los datos para la construcción de las Funciones de Información de los Tests (FITs) y los errores estándares de medida resultantes tras la aplicación de cada uno de los modelos también fue suministrada por MULTILOG.

3. Resultados

3.1. Supuesto de Unidimensionalidad

El Análisis de Componentes Principales (prueba de esfericidad de Bartlett, $\chi^2 = 6419.9$, $p < .0001$, $KMO = .91$) mostró una solución con un único componente dominante cuyo autovalor fue de 4.4 y que explicó el 54.8% de la varianza. El cociente entre el autovalor del primer y segundo componente resultó de 6.03. Las cargas factoriales de los ítems oscilaron entre .63 y .83 (Tabla 1). Estos resultados corroboran la unidimensionalidad de los datos considerando como criterios: a) un porcentaje de varianza explicada por el primer componente mayor a 40% (Carmines & Zeller, 1979) y b) un cociente entre el autovalor del primer componente y del segundo mayor a 5 (Martínez Arias et al., 2006).

Tabla 1.

Análisis de Componentes Principales

Ítem	Componentes							
	1	2	3	4	5	6	7	8
1	.77	-.30	-.25	-.01	.32	-.12	.21	-.31
2	.70	-.01	.52	.26	.30	.29	.00	.06
3	.69	.41	-.03	-.52	.11	.10	.14	.11
4	.75	-.34	-.17	-.02	-.39	.36	.11	.04
5	.83	.08	.04	-.12	-.09	-.01	-.47	-.25
6	.80	-.24	-.21	.04	.13	-.22	-.17	.39
7	.74	-.01	.44	.00	-.31	-.34	.19	-.02
8	.63	.05	-.32	.44	-.08	-.01	.07	-.01
Autovalor	4.38	.73	.71	.58	.48	.42	.37	.33
Varianza descripta (%)	54.8	9.1	8.8	7.3	6.1	5.2	4.6	4.2

3.2. Estimación y Ajuste

El proceso de estimación de los parámetros de los ítems finalizó en el ciclo 34 para el MRG y en el ciclo 28 para el MCP con un criterio de convergencia de 0.001. En la Tabla 2 se exhiben los valores obtenidos en este proceso para ambos modelos y las medias y desviaciones típicas correspondientes. Las proporciones observadas y

esperadas de las categorías para ambos modelos mostraron residuos inferiores a 0.001 por lo que, según este dato proporcionado por MULTILOG, los modelos tuvieron un ajuste satisfactorio a los datos.

En la misma línea, las estimaciones alcanzaron el criterio de convergencia empleando una cantidad de iteraciones razonable y los errores de estimación resultaron bajos. Los parámetros

presentaron, en términos generales, valores dentro de un rango esperable. La escala adoptada tanto para θ como para los parámetros de umbral es arbitraria y no presenta unidades de medida intrínsecamente definidas. Sin embargo, se suele definir por consenso su origen según la escala estandarizada con media 0 y desviación estándar 1. Así como es común encontrar que la mayoría de los individuos se ubiquen a menos de 3 desviaciones por encima y por debajo de la media de θ , también es razonable esperar que los b y los β oscilen entre -3 y 3. Sólo el parámetro $b_7 = -3.16$ del ítem 8 correspondiente a la aplicación del MRG excedió el rango esperable pero no se considera una desviación importante como para desestimar la calidad del ítem de manera global. Se obtuvieron evidencias empíricas de la invarianza de las estimaciones del rasgo latente respecto de cambios en la composición del test. La correlación entre los θ estimados con dos subconjuntos de cuatro ítems

seleccionados al azar fue de .85 ($p < .001$) para el MRG y de .78 ($p < .001$) para el MCP. Además se corroboró la invarianza de las propiedades de los ítems estimadas a partir de dos submuestras de participantes distribuidos de forma aleatoria ($n_1 = 938$ y $n_2 = 937$). En el MCP la asociación que resultó más elevada fue entre los valores de los parámetros β_2 estimados para cada submuestra ($r = .99$; $p < .001$) mientras que las más bajas se hallaron entre los correspondientes de β_4 y β_5 ; estos últimos con correlaciones de .92 ($p < .001$). Las asociaciones entre las estimaciones obtenidas para cada submuestra fueron ligeramente más elevadas para el MRG. Los parámetros b_1 , b_2 , b_3 presentaron correlaciones de .99 ($p < .001$) al calcular la asociación entre sus correspondientes valores estimados cada una de las submuestras. En cambio, las correlaciones más bajas de este modelo se obtuvieron para los parámetros a ($r = .95$; $p < .001$) y b_5 ($r = .95$; $p < .001$).

Tabla 2.

Parámetros estimados para los modelos Respuesta Graduada y Crédito Parcial.

Ítem	MRG						MCP				
	a	b_1	b_2	b_3	b_4	b_5	β_1	β_2	β_3	β_4	β_5
1	1.29	-1.60	-1.00	-0.38	0.35	1.25	-1.07	-0.99	-0.51	0.23	1.11
2	0.97	-2.07	-1.13	-0.26	0.74	1.78	-1.58	-1.02	-0.38	0.69	1.47
3	0.99	-2.85	-2.02	-0.99	0.31	1.58	-2.09	-2.01	-1.17	0.32	1.43
4	1.22	-0.87	-0.29	0.23	0.95	1.86	-0.26	-0.18	-0.01	0.88	1.75
5	1.66	-1.85	-1.07	-0.32	0.62	1.60	-1.83	-1.21	-0.54	0.75	1.79
6	1.55	-1.73	-1.12	-0.43	0.48	1.42	-1.44	-1.26	-0.69	0.53	1.46
7	1.18	-2.32	-1.54	-0.64	0.55	1.73	-1.89	-1.62	-0.87	0.63	1.76
8	0.86	-3.16	-1.97	-0.76	0.63	1.94	-2.44	-1.81	-0.90	0.64	1.64
M	1.21	-2.06	-1.27	-0.44	0.58	1.64	-1.57	-1.26	-0.63	0.59	1.55
DE	0.28	0.68	0.53	0.34	0.20	0.22	0.63	0.53	0.34	0.20	0.22

Nota: MRG = Modelo de Respuesta Graduada; MCP = Modelo de Crédito Parcial; M = Media de los parámetros estimados para los ítems; DE = Desviación estándar de los parámetros estimados para los ítems.

3.3. Parámetros

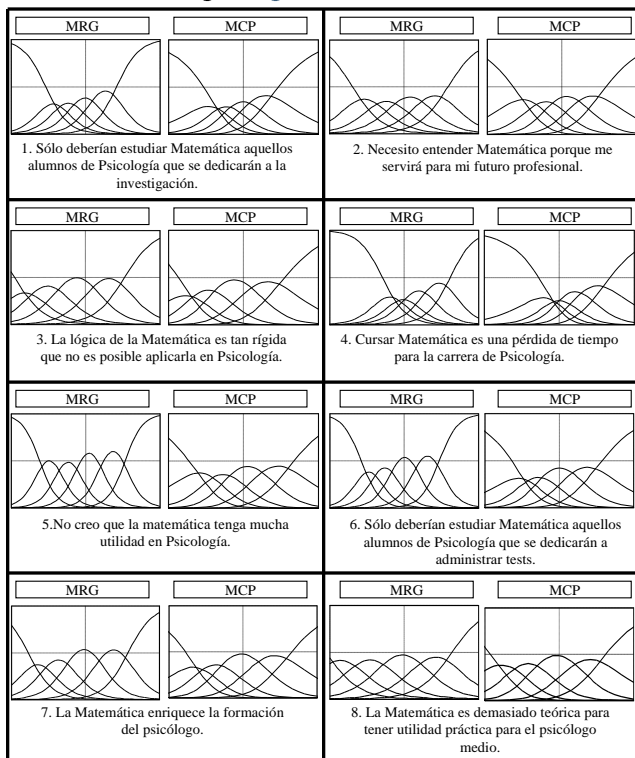
Los parámetros a estimados para el MRG tuvieron un valor promedio moderado ($a_{promedio} = 1.21$) y una limitada variabilidad ($a_{DE} = 0.28$). Estos resultados revelan que la escala presenta una capacidad discriminativa media y que parece haber cierta homogeneidad en la capacidad discriminativa de los ítems. Esto último justifica la conveniencia de ensayar un modelo más parsimonioso como es el MCP en virtud de que, justamente, tiene como requerimiento igual discriminación para todos los ítems. A pesar de que el MCP supone una discriminación de 1 para los ítems, el programa MULTILOG permite estimar

libremente un parámetro de discriminación para este modelo bajo la condición de que sea constante. El valor obtenido en esta estimación fue de 1.03 ($SE = 0.01$), lo cual indica que la capacidad discriminativa de los ítems no subestima ni sobreestima la esperada por el modelo.

Los parámetros b del MRG estuvieron comprendidos entre -3.16 y 1.94 mientras que los β del MCP se ubicaron en un rango más estrecho del rasgo desde -2.44 a 1.79. Los parámetros que se encuentran en la posición central (b_3 y β_3) adoptaron en su gran mayoría valores negativos, por lo que se observa clara una tendencia a estar localizados en

niveles medio-bajos de Utilidad. Estos resultados indican, contrario a lo podría presuponerse, que para los estudiantes de Psicología resulta “fácil” estar de acuerdo con las sentencias que proponen los enunciados de los ítems (o estar en desacuerdo para los ítems redactados en sentido inverso).

En ambos modelos todas las categorías de respuesta de los ocho ítems llegan a ser máximamente probables en algún tramo del recorrido del rasgo (Figura 1).



Nota: MRG = Modelo de Respuesta Graduada; MCP = Modelo de Crédito Parcial.

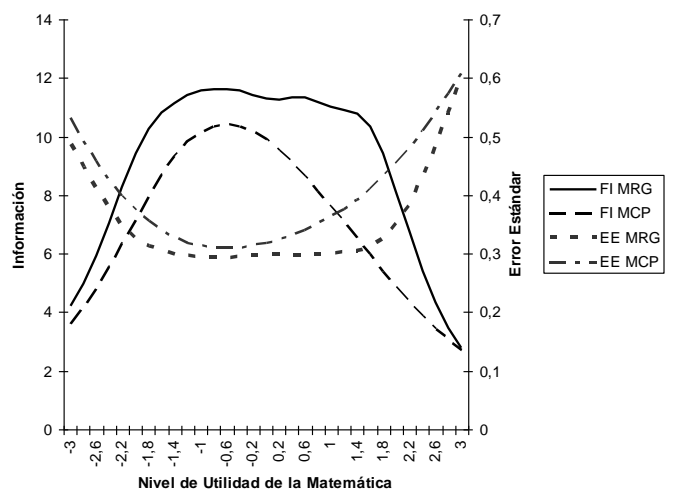
Figura 1. Curvas Características de las Categorías de Respuesta de los ítems.

No obstante, este tramo es extremadamente corto para la segunda categoría de los ítems 1, 3 y 4. Si se observa la aplicación del MRG se puede apreciar que para estos ítems los parámetros b_1 y b_2 se localizaron próximos entre sí. La situación es más crítica en la modelización con el MCP porque, en general, los parámetros de localización presentaron una menor amplitud y, sumado a esto, la discriminación de los ítems resultó más baja. Como consecuencia, además de los ítems mencionados para el MRG, también presentaron el mismo problema los ítems 6 y 7 en la aplicación del MCP.

3.4. Funciones de Información

La Figura 2 permite comparar las FITs y los errores estándares en función del rasgo latente obtenidos para la escala de Utilidad con los MRG y MCP.

Aunque no se ha definido en la bibliografía psicométrica un criterio para determinar cuándo una FIT es elevada, la comparación de las funciones obtenidas para los modelos aportan una medida de base que orienta el análisis. El MRG mostró un valor máximo de información de 11.6 mientras que el MCP sólo llegó a 10.6. Consecuentemente, dada la relación inversa, el MRG tuvo error estándar más bajo de 0.29 mientras que el del MCP fue de 0.31. Ambos modelos alcanzaron su máxima información y menor error para un nivel de $\theta = -0.6$. La FIT del MRG fue relativamente constante entre los niveles -1.8 y 1.8 del rasgo latente. Por ende, se observa que resultó elevada incluso para los niveles más extremos de la variable, ventaja que le permite destacarse por sobre la FIT del MCP.



Nota: FI MRG = Función de Información del Modelo de Respuesta Graduada; FI MCP = Función de Información del Modelo de Crédito Parcial; EE MRG = Error Estándar del Modelo de Respuesta Graduada; EE MCP = Error Estándar del Modelo de Crédito Parcial.

Figura 2. Función de Información del Test y Error Estándar de medida a partir de la modelización con el Modelo de Respuesta Graduada y el Modelo de Crédito Parcial.

4. Discusión

El uso de la TRI para el análisis de tests de comportamiento típico es por demás infrecuente (Abal, Lozzia, Aguerri, Galibert & Attorresi, 2010a; Reise & Waller, 2003, 2009). Su aplicación brinda información para profundizar el estudio de la estructura interna de una prueba mediante un análisis pormenorizado de los ítems (Simms, 2008). En el

marco de esta investigación, la TRI ha mostrado su utilidad para establecer una relación entre los niveles de la variable estudiada y la tendencia al acuerdo de los estudiantes con los contenidos de cada ítem. Los ocho ítems de la escala mostraron un adecuado ajuste tanto al MRG como al MCP según el estudio de los residuos entre las proporciones observadas y esperadas que ofrece MULTILOG. También resultaron favorables las evidencias obtenidas mediante indicadores indirectos del ajuste surgidos del proceso de estimación (cantidad de ciclos, valores estimados y errores de estimación). Con respecto a la invarianza de los parámetros de los ítems y de los individuos el MRG tuvo mejores resultados. El MCP presentó menores correlaciones de los parámetros de cada ítem estimados a partir de diferentes muestras y de los parámetros de los estudiantes con dos subconjuntos diferentes de ítems de la prueba.

Resulta conveniente resaltar nuevamente que los índices aplicados para la evaluación del ajuste pueden ser considerados débiles (Revuelta et al., 2006). Una evaluación más exhaustiva del ajuste debería analizar la discrepancia entre proporciones observadas y esperadas de cada categoría en cada nivel del rasgo latente (Abad, Olea, Ponsoda & García, 2011). Sin embargo, la relativa novedad de los modelos politómicos origina que los métodos de evaluación de ajuste de los datos al modelo aún se encuentren en fase de exploración y redefinición (Sueiro & Abad, 2009). Tampoco se han considerado indicadores de ajuste de los sujetos al modelo porque el objetivo de esta investigación es conocer los parámetros de los ítems. Como afirmaron Revuelta et al (2006), el estudio de ajuste de los sujetos debe realizarse en etapas posteriores, cuando se efectúa la aplicación del test, y no durante la fase de calibración de los ítems. Bajo estas circunstancias, es posible suponer que el número de individuos desajustados es pequeño en comparación con el total de la muestra y, por ende, no afectarán al ajuste del modelo.

La FIT del MRG presentó valores más elevados que la del MCP para todo nivel del rasgo latente, lo que refleja que su modelización consigue mediciones más precisas. El error de medida según el MRG fue más bajo y parejo en gran parte del recorrido del rasgo. Esto implica que las estimaciones de los diferentes niveles de θ se llevan a cabo con un error estándar similar. La capacidad discriminativa de los

ítems se asocia a la precisión de una prueba (Abad et al., 2011). Es por esta razón que la diferencia a favor del MRG en los máximos alcanzados por las FITs de los modelos puede justificarse en parte por la condición de homogeneidad de los parámetros de discriminación que requiere el MCP.

Otro aspecto que afecta a la forma de las FIT son los rangos de variación de los parámetros de umbral. Las distancias halladas entre los diferentes parámetros de umbral del MRG fueron considerables permitiendo que la información aportada por los ítems se mantenga constante en un rango extenso de la variable. En cambio, los parámetros β del MCP mostraron una escasa variabilidad principalmente para las categorías inferiores. Como demostraron Dodd y Koch (1987), en la FIT del MCP la información se acumula en niveles del rasgo cercanos a los parámetros de umbral. Como consecuencia, cuando los valores de los parámetros β se localizan próximos entre sí la FIT tiende a un mayor grado de apuntamiento. En este caso, se puede observar que la FIT del MCP rinde su máxima información en torno a niveles θ que se encuentran a 0.5 desviaciones por debajo de la media del rasgo.

La comparación de las aplicaciones de los modelos politómicos en relación con el ajuste y la precisión de las mediciones permiten afirmar que el MRG ofrece una estructura paramétrica más adecuada para representar las respuestas de los estudiantes a los ítems de la escala de Utilidad de la Matemática. Este resultado está en línea con hallazgos de otros autores, quienes ya habían mostrado que la aplicación del MRG mostraba ventajas por sobre otros modelos politómicos y dicotómicos de la TRI (Asún & Zúñiga, 2008; Baker, Rounds, & Zevon, 2000).

Cabe señalar que parte de las ventajas registradas en la modelización con MRG se asientan en la inclusión del parámetro de discriminación de los ítems. Sin embargo, se ha mencionado anteriormente que la escasa variabilidad de los a del MRG podría reflejar cierta homogeneidad en la capacidad discriminativa de los ítems. En este sentido, es importante preguntarse si la ganancia obtenida gracias al parámetro a es suficientemente grande como para compensar la pérdida de parsimonia. La respuesta a esta cuestión no es sencilla porque implica sopesar dos aspectos altamente valorados: la precisión del instrumento y la parsimonia de un

modelo. El beneficio sustancial que aporta la modelización con MRG es que permite una medición más precisa que el MCP. La función del error de estándar del MRG muestra que error se es prácticamente constante en gran parte del recorrido del rasgo, por lo que la mayoría de las personas son medidas con similar precisión sin importar cuál sea su nivel de rasgo. En cambio, el escaso rango de variación de los β del MCP genera que la precisión sea máxima sólo para los valores medio-bajos del rasgo, mientras que el resto de los niveles pueden verse perjudicados.

En cuanto a las limitaciones del presente trabajo debe reconocerse que sólo se ha centrado en recolección de evidencias de validez basadas en la estructura interna de la escala. Como afirman Maydeu-Olivares y García-Forero (2010), el estudio del ajuste de un modelo no es por sí solo una garantía de su utilidad. Las creencias sobre la utilidad de la Matemática juegan un papel importante como fuerzas que favorecen u obstaculizan la disposición al estudio de los alumnos (Gómez Chacón, 2005). Una valoración negativa de la utilidad de la Matemática campo de la Psicología no sólo podría incidir en la cabal comprensión de los contenidos propios de esta materia y otras afines sino que también limitaría los posibles campos de especialización del futuro desarrollo profesional del estudiante. Futuras investigaciones tendrán como objetivo la obtención de evidencias de validez basadas sobre fuentes externas que permitan establecer las relaciones del rasgo evaluado con otros constructos o variables observables. Ulteriores estudios también se centrarán en realizar las adaptaciones correspondientes a los ítems para aplicar la escala en estudiantes de otras carreras humanístico-sociales. Un aumento en la heterogeneidad de la muestra permitiría obtener indicadores para estudiar la invariancia de los parámetros estimados en las distintas poblaciones.

Agradecimientos

Subsidios de la Universidad de Buenos Aires (UBACyT N° 20020100100346), de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 2011 N° 0826) y de la Beca Postdoctoral otorgada al Dr. Facundo Juan Pablo Abal por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Referencias

- Abad, F., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Abal, F. (2013). *Comparación de modelos dicotómicos y politómicos de la Teoría de Respuesta al Ítem aplicados a un test de comportamiento típico*. Tesis Doctoral no publicada. Universidad de Buenos Aires, Argentina.
- Abal, F., Lozzia, G., Aguerri, M., Galibert, M. & Attorresi, H. (2010a). La escasa aplicación de la Teoría de Respuesta al Ítem en Tests de Ejecución Típica. *Revista Colombiana de Psicología*, 19(1) 111-122.
- Abal, F., Lozzia, L., Aguerri, M., Galibert, M. & Attorresi, F. (2010b). *La Utilidad de la Matemática en el campo de la Psicología desde la perspectiva de los estudiantes*. Trabajo presentado en el II Encuentro de Docentes e Investigadores de Estadística en Psicología, Buenos Aires, Argentina.
- Álvarez, Y. & Ruiz, M. (2010). Actitudes hacia las matemáticas en estudiantes de ingeniería en universidades autónomas venezolanas. *Revista de Pedagogía*, 31 (89), pp. 225 – 249.
- Asún, R. & Zúñiga, C. (2008). Ventajas de los Modelos Politómicos de la Teoría de Respuesta al Ítem en la Medición de Actitudes Sociales. El Análisis de un Caso. *Psykhe*, 17, 103 - 115.
- Auzmendi, E. (1992). *Las actitudes hacia la matemática-estadística en las enseñanzas medias y universitarias*. Bilbao: Mensajero.
- Baker, J. G., Rounds, J. B. & Zevon, M. A. (2000). A Comparison of Graded Response and Rasch Partial Credit Models with Subjective Well-Being. *Journal of Educational and Behavioral Statistics*, 25, 253-270.
- Bazán J. & Sotero H. (1998). Una aplicación al estudio de actitudes hacia la Matemática en la UNALM. *Anales Científicos de la Universidad Nacional Agraria La Molina*, 36, 60-72.
- Bermejo, V. (1996). Enseñar a comprender las matemáticas. En J. Beltrán & C. Genovard (Eds.), *Psicología de la Instrucción I*. (pp. 256-279). Madrid: Síntesis.
- Carmines, E. & Zeller, R. (1979). *Reliability and validity assessment*. Londres: Sage.
- Coolican, H. (1994). *Métodos de investigación y estadística en psicología*. México: El Manual Moderno.
- Cortada, N., Rodríguez Feijóo, N. & Kohan Cortada, A. (2008). *Análisis de las sugerencias para modificar las actitudes negativas hacia la estadística*. Trabajo presentado en el I Encuentro de Docentes e Investigadores de Estadística en Psicología, Buenos Aires, Argentina.
- Dodd, B. & Koch, W. (1987). Effects of Variations in Item Step Values on Item and Test Information in the Partial

- Credit Model. *Applied Psychological Measurement*, 11, 371-384.
- 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 males and females. *Catalog of Selected Documents in Psychology*, 6, 31.
- Gempp, R., Denegri, M., Caprile, C., Cortés, L., Quesada, M. & Sepúlveda, J. (2006). Medición de la Alfabetización Económica en Niños: Oportunidades Diagnósticas con el Modelo de Crédito Parcial. *Psykhé*, 15(1), 13-27.
- Gil, J., García, E., Rodríguez, G., Guijarro, O. & López, M. A. (1995). Utilidad de la Estadística desde la perspectiva de los alumnos de Ciencias de la Educación. *Enseñanza*, 13, 27-41.
- Gil, N., Blanco, L. & Guerrero, E. (2006). El dominio afectivo en el aprendizaje de las matemáticas. *Revista electrónica de Investigación Psicoeducativa*, 8, 47-72.
- Goldin, G. A., Roesken, B. & Toerner, G. (2009). Beliefs – No longer a hidden variable in mathematics teaching and learning processes. En J. Maass & W. Schöglmann (Eds.), *Beliefs and Attitudes in Mathematics Education: New Research Results* (pp. 1-18). Rotterdam: Sense Publishers.
- Gómez Chacón, I. (2005). *Matemática emocional. Los afectos en el aprendizaje matemático*. España: Narcea.
- Gómez Chacón, I. (2010). Tendencias actuales en investigación en matemáticas y afecto. En M. M. Moreno, A. Estrada, J. Carrillo, & T. A. Sierra, (Eds.), *Investigación en Educación Matemática XIV* (pp. 121-140). Lleida: SEIEM.
- Gray-Little, B., Williams, V. S. L. & Hancock, T. D. (1997). An Item Response Theory Analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 23, 443-451.
- Hernández, G. (2011). Estado del arte de creencias y actitudes hacia las matemáticas. *Cuadernos de Educación y Desarrollo*, 3 (24). Extraído el 10 de abril de 2011 desde <http://www.eumed.net/rev/ced/24/ghs.htm>
- Kang, T. & Cohen, A. S. (2007). IRT Model Selection Methods for Dichotomous Items. *Applied Psychological Measurement*, 31, 331-358.
- Martínez Arias, M. R., Hernández Lloreda, M. V. & Hernández Lloreda, M. J. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Martínez, G. A. & Gaitán, M. M. (2013). *Sistemas de creencias y rendimiento en Matemática en estudiantes de Ingeniería*. Trabajo presentado en el VII Congreso Iberoamericano de Educación Matemática, Montevideo, Uruguay.
- Martínez, O. J. (2008). Actitudes hacia la matemática. *Sapiens. Revista Universitaria de Investigación*, 9 (1), 237-256.
- Masters, G. N. & Wright, B. D. (1997). The Partial Credit Model. En W. J. Van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, (pp. 101-121). New York: Springer.
- Maydeu-Olivares, A. & García-Forero, C. (2010). Goodness of fit testing. En P. Peterson, E. Baker & B. McGaw, B. (Eds.), *International Encyclopedia of Education (3rd ed.)* (pp. 190-196). Oxford: Elsevier.
- McLeod, D. & McLeod, S. (2002). Synthesis – Beliefs and Mathematics Education: Implications for Learning, Teaching and Research. En G. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 115-126). Dordrecht: Kluwer Academic Publishers.
- Montero, I. & León, O. G. (2005). Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115-127.
- Muñiz, J., Fidalgo, A. M., García Cueto, E., Martínez, R. & Moreno, R. (2005). *Análisis de los ítems*. Madrid: La Muralla.
- Palacios, A., Arias, V. & Arias, B. (2014). Attitudes Towards Mathematics: Construction and Validation of a Measurement Instrument. *Revista de Psicodidáctica*, 19(1), 67-91.
- Ponsoda, V. (1990). Un punto de vista sobre la docencia de la Estadística en Psicología. *Estadística Española*, 31, 481-489.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute for Educational Research.
- Reise, S. & Waller, N. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods*, 8, 164–184.
- Reise, S. & Waller, N. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27-48.
- Revuelta, J., Abad, F. J. & Ponsoda, V. (2006). *Modelos Políticos de respuesta al ítem*. Madrid: La Muralla.
- Rodríguez Feijóo, N. (2011). Actitudes de los estudiantes universitarios hacia la estadística. *Interdisciplinaria*, 28 (2), 199-205.
- Rubio, V. J., Aguado, D., Hontangas, P. M. & Hernández, J. M. (2007). Psychometric Properties of an Emotional Adjustment Measure. An Application of the Graded Response Model. *European Journal of Psychological Assessment*, 23 (1), 39-46.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 17.

- Simms, L. (2008). Classical and Modern Methods of Psychological Scale Construction. *Social and Personality Psychology Compass*, 2, 414 – 433.
- Sueiro, M. J. & Abad, F. J. (2009). Bondad de ajuste en ítems politómicos: tasas de error de tipo I y potencia de tres índices de ajuste. *Psicothema*, 21(4), 639-645.
- Tapia, M. & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8, 16-21.
- Thissen, D. (2003). *MULTILOG*. Chicago: Scientific Software International.