

Los modelos de mezcla de la Teoría de respuesta al ítem

The Item Response Theory Mixture Models

Armél Brizuela-Rodríguez¹
Universidad de Costa Rica, Costa Rica

Resumen. Se presentan los modelos de mezcla de la Teoría de respuesta al ítem y cómo estos pueden ser utilizados para identificar subagrupaciones no observadas de examinados denominadas como clases latentes. Asimismo se ejemplifica la utilidad del modelo de mezcla de la Teoría de respuesta al ítem de dos parámetros, mediante el cual se logra detectar la presencia de ítems en una escala de depresión que miden de manera distinta a los examinados en comparación con el resto de ítems que componen dicho instrumento. Finalmente, se presentan algunas recomendaciones generales en cuanto a la complementariedad que debe existir entre la teoría sustantiva que subyace al desarrollo de una escala y la aplicación de los modelos de mezcla de la Teoría de respuesta al ítem.

Palabras clave. Psicometría, Teoría de respuesta al ítem, modelos de mezcla, clases latentes, Escala de depresión geriátrica.

Abstract. The Item Response Theory mixture models and how these can be used to identify unobserved sub-groups of examinees, known as latent classes, are presented. The usefulness of the two-parameter mixture model parameters is exemplified by detecting the presence of items in a depression scale that measure the examinees differently in comparison with the rest of items. Finally, some general recommendations regarding the complementarity that should exist between the substantive theory underlying the development of a scale and the use of these models are given.

Keywords. Psychometrics, Item Response Theory, mixture models, latent classes, Geriatric Depression Scale.

¹Armél Brizuela-Rodríguez. Programa Prueba de Aptitud Académica, Instituto de Investigaciones Psicológicas, Universidad de Costa Rica. Dirección postal: 11501-2060, San José, Costa Rica. Email: armel.brizuelarodriguez@ucr.ac.cr



Introducción

La medición psicoeducativa cumple un rol fundamental en la actualidad en el campo de las ciencias sociales y de la salud, en las cuales se suele conceptualizar como la asignación sistemática de números a personas como una forma de representar características psicológicas, tales como inteligencia, personalidad, motivación, aptitud, actitudes, calidad de vida, entre otros (Raykov & Marcoulides, 2011). A diferencia de muchos atributos físicos (peso, edad, estatura, color de piel, etc.), las variables de interés en estos campos del saber no pueden ser observadas directamente por lo que instrumentos para cuantificar este tipo de variables se consideran como indicadores indirectos de una o varias variables no observadas (constructos o variables latentes).

Un constructo o variable latente es una fuente oculta de varianza (o de covarianza) en un conjunto de puntuaciones observadas, las cuales se obtienen a partir de la aplicación de un test a un conjunto de personas (estudiantes, pacientes, aspirantes a ingreso a una institución, etc.). En este sentido, un test es un dispositivo evaluativo o procedimiento mediante el cual se obtiene una muestra de la conducta de los examinados en un dominio especificado, el cual es posteriormente evaluado y calificado mediante un proceso estandarizado (AERA, APA & NCME, 2014). De esta manera, los puntajes derivados de la aplicación de un test son indicadores de un constructo, el cual se considera como la causa principal de su variabilidad (Borsboom, Mellenbergh & Heerden, 2004).

Para medir este tipo de variables, es necesario acudir a técnicas especializadas que permitan cuantificar el nivel de las personas en un determinado constructo o la pertenencia de estas a diferentes clases latentes. Para ello, la psicometría se ha consolidado como una disciplina científica que estudia los problemas inherentes a la medición de variables latentes y que provee de modelos para abordarlos. Ahora bien, uno de los pasos cruciales en este proceso es establecer si la variable latente por medir (ansiedad, depresión, conocimientos, etc.) es categórica, continua o una combinación de ambas (De Ayala, 2009). En el caso

de que sea categórica, los sujetos serán clasificados en grupos cualitativamente diferentes denominados *clases latentes*. Por otra parte, si el constructo de interés se considera una variable continua, entonces cada persona presentará un determinado nivel en un continuo. Una tercera posibilidad es que el espacio latente se considere una combinación entre variables categóricas y continuas, de modo que cada persona sería asignada a una clase latente y además presentará un nivel en el constructo; en este último escenario es necesario acudir a los modelos de mezcla.

Por otra parte, las poblaciones de interés en las ciencias sociales y de la salud suelen ser muy heterogéneas y es frecuente que las fuentes de dicha heterogeneidad no sean conocidas a priori (Lubke & Muthen, 2005). Por ejemplo, los examinados pueden utilizar diferentes estrategias cuando contestan ítems pertenecientes a un test, de modo que algunas de ellas serán apropiadas para resolver los ítems y otras no (Von Davier, 2010). Así las cosas, en principio es posible observar que en una misma muestra los examinados conformen distintas subagrupaciones en función de las estrategias utilizadas para contestar los ítems. Ante esta situación, sería conveniente emplear un modelo de mezcla para analizar las respuestas a los ítems, por cuanto es plausible suponer que los parámetros de los ítems y de los examinados son distintos en cada una de las posibles subagrupaciones no observadas que componen la población en estudio (Vermunt, 2014).

Cuando las fuentes de heterogeneidad son conocidas (sexo, edad, zona de residencia, etc.) se pueden utilizar modelos multigrupo como el análisis del funcionamiento diferencial del ítem para comparar los parámetros estimados en los grupos conformados a partir de las fuentes de variabilidad irrelevante al constructo que se pretende medir (Raykov, Marcoulides, Lee & Chang, 2013). No obstante, los grupos podrían conformarse a partir de variables no observadas, por lo cual la pertenencia de los examinados a las posibles subagrupaciones únicamente puede ser inferida a partir de los datos. Los modelos en los que se asume que la fuente de heterogeneidad no es observable están diseñados para detectar conglomerados de sujetos

con patrones de respuesta similares en un conjunto observado de variables, como por ejemplo las respuestas a un test. De hecho, el uso de los modelos de mezcla de la Teoría de respuesta al ítem ha resultado ser de gran utilidad para complementar los análisis del funcionamiento diferencial de los ítems (DeMars & Lau, 2011; Lee & Beretvas, 2014).

Cabe señalar que la mayoría (si no es que todos) de los modelos de mezcla para analizar datos provenientes de la aplicación de test pueden ser vistos como extensiones del modelo básico utilizado en el análisis de clases latentes, el cual es un modelo de mezcla debido a que para estimar la probabilidad de observar un determinado vector de respuestas a un conjunto de ítems es necesario multiplicar entre sí las probabilidades de respuesta correcta en cada clase latente y posteriormente sumar dichos productos. Por ejemplo, si se contara con 10 ítems dicotómicos en principio podrían observarse 1024 patrones de respuesta diferentes y sería posible agrupar a los examinados de acuerdo con estos de tal manera que quienes presenten patrones similares pertenezcan a un mismo grupo.

En el contexto del análisis de clases latentes es posible agrupar dichos patrones de respuesta (y, por ende, a los sujetos que muestran dichos patrones) en un número reducido de k clases latentes ($k < 1024$) de modo que los patrones de respuesta para los examinados pertenecientes a una misma clase son más similares entre sí que con respecto a los patrones de quienes pertenezcan a otra clase latente.

Un supuesto fundamental en el análisis tradicional de clases latentes es el de la independencia local, el cual se refiere a que las variables observadas no covarían entre sí a lo interno de cada clase. De este modo, asumiendo la independencia local de los ítems a lo interno de las clases latentes, la probabilidad condicional de un determinado patrón de respuestas es igual al producto de las probabilidades individuales de respuesta de ese patrón (Collins & Lanza, 2010).

Los modelos de mezcla de la Teoría de respuesta al ítem son aquellos que permiten no solo estimar

la pertenencia de los examinados a alguna de las posibles clases latentes, sino también su nivel en el constructo que se pretende medir. Dado que estos modelos incorporan variables latentes continuas y categóricas, son especialmente útiles para analizar datos generados a partir de la medición de un constructo continuo en una muestra compuesta por dos o más subagrupaciones no observadas. En este sentido, son modelos de clases latentes en los que las probabilidades de observar una determinada respuesta en una clase latente se modelan de acuerdo con alguno de los modelos tradicionales de Teoría de respuesta al ítem (Formann & Kohlmann, 2002).

Lo anterior implica que en estos modelos se flexibiliza el supuesto de independencia local entre los ítems a lo interno de cada clase, para lo cual se plantea la existencia de una o varias variables latentes continuas que explican las dependencias entre los ítems en cada una de las clases (Sterba, 2013). A diferencia del análisis tradicional de clases latentes, esto permite plantear la existencia de diferencias individuales sistemáticas entre los individuos que pertenezcan a una misma clase latente.

En el modelo de la Teoría de respuesta al ítem de tres parámetros para ítems dicotómicos (Hambleton, Swaminathan & Rogers, 1991) se plantea que la probabilidad de contestar correctamente un ítem (o de elegir la opción codificada con el mayor valor) depende de su dificultad (b), de su discriminación (a) y del azar (c), así como del nivel en el constructo del examinado (θ):

$$P(I=1|a,b,c,\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta_i - b_i)}}{1 + e^{a_i(\theta_i - b_i)}} \quad (1)$$

Cuando este modelo se combina con el análisis de clases latentes, entonces es posible estimar la probabilidad de elegir la opción de mayor valor para cada una de las g clases latentes así como la proporción de examinados que pertenecen a cada clase (π_g), a saber:

$$P(I = g | a, b, c, \theta, g) = \sum_{g=1}^g \pi_g \left(c_{ig} + (1 - c_{ig}) \frac{e^{a_{ig}(\theta_g - b_{ig})}}{1 + e^{a_{ig}(\theta_g - b_{ig})}} \right) \quad (2)$$

En los modelos de mezcla de la Teoría de respuesta al ítem tanto la distribución de la habilidad como las probabilidades de respuesta dependen de la pertenencia a una determinada clase latente. Dado que se asume la existencia de dos o más clases a las que podrían pertenecer los examinados, el número de estas debe ser inferido utilizando estrategias que implican la clasificación de los patrones de respuesta observados (Von Davier & Yamamoto, 2007). Así pues, estos modelos asumen una estructura más compleja a lo interno de cada clase en comparación con el análisis de clases latentes. Específicamente, además de incluir una variable latente categórica de clasificación también incluyen una continua para modelar el nivel en el constructo de los examinados.

Cabe señalar que, como se puede observar en la Fórmula 2, el modelo utiliza una transformación logito como función de enlace al modelo, por lo cual se asume que la distribución de probabilidad condicional de la respuesta al ítem es binomial (Bernoulli). Además, para que el modelo sea estimable el resultado de sumar los π_g de cada clase latente debe ser igual a 1 y la suma de los b_{ig} en cada clase debe ser igual a 0 (Rost, 1990).

Los parámetros de los ítems y los de las personas podrían ser diferentes en cada clase. Por lo tanto, estos modelos se utilizan cuando se sospecha que cada examinado podría pertenecer a una de varias posibles clases exhaustivas y mutuamente excluyentes que corresponden a las posibles estrategias para contestar los ítems (Mislevy & Verhelst, 1990). Dada su gran flexibilidad, han sido estudiados para conocer sus fortalezas, debilidades y aplicaciones.

Desde una perspectiva teórico-metodológica, ha habido un gran interés por investigar aspectos relacionados con los métodos de estimación de los parámetros, la consistencia en la asignación de los examinados a las clases latentes, los índices de ajuste utilizados para seleccionar los modelos, la comparabilidad de los parámetros estimados en cada

clase y la utilidad de incluir covariables como parte de la caracterización de los examinados en las clases latentes (Li, Cohen, Kim & Cho, 2009; Muthén & Lubke, 2007; Nylund, Asparouhov & Muthén, 2007; Paek & Cho, 2015; Preinerstorfer & Formann, 2011; Smit, Kelderman & van der Flier, 2000; Smit, Kelderman & van der Flier, 1999; Tueller y Lubke, 2010; Von Davier & Molenaar, 2003; Willse, 2011). En esta línea de investigación, recientemente se ha comenzado a explorar la sensibilidad de estos modelos en la detección de clases latentes espúreas (Alexeev, Templin & Cohen, 2011; Chen & Jiao, 2013), la utilidad de incorporar estructuras multinivel en los análisis (Cho & Cohen, 2010), la estimación mediante métodos bayesianos y la inclusión de covariables observadas (Dai, 2013).

Finalmente, a la fecha se han explorado una gran diversidad de aplicaciones de estos modelos, como la generación de evidencias de validez referidas a la invarianza de los parámetros de los ítems (Baghaei & Carstensen, 2013) y a la dimensionalidad de los test (Hong & Min, 2007), la identificación de posibles fuentes del funcionamiento diferencial de los ítems (Choi, Alexeev & Cohen, 2014; Cohen & Bolt, 2005; DeMars & Lau, 2011; Elosúa & López, 2006; Frick, Strobl & Zeileis, en prensa; Meij, Kelderman & Flier, 2010; Oliveri, Ercikan, Zumbo & Lawless, 2014; Tay, Newman & Vermunt, 2011), el estudio de las diferencias individuales relacionadas con la tendencia a elegir ciertas opciones de respuestas (Bolt, Cohen & Wollack, 2001; Carter, Dalal, Lake, Lin & Zickar, 2011; Egberink, Meijer & Veldkamp, 2010; Eid & Rauber, 2000; Meij, Kelderman & Flier, 2008; Meiser & Machunsky, 2008; Mneimneh, Heeringa, Tourangeau & Elliott, 2014), la identificación de la existencia de tipos diferentes de examinados que utilizan estrategias distintas de razonamiento (De Boeck & Rijmen, 2003; Embretson, 2007), la identificación de subagrupaciones de examinados en función de la velocidad de estos para contestar los ítems ubicados al final de un test (Bolt, Cohen & Wollack, 2002; Meyer, 2010), la detección de examinados con bajos niveles de motivación para contestar test de bajas consecuencias (Mittelhaeuser, Béguin & Sijtsma, 2013), la detección de diferentes estilos de respuesta respecto de la expresión de la ira

(Gollwitzer, Eid & Jürgensen, 2005), la identificación de subpoblaciones de adolescentes que incurrir en el consumo de drogas y conductas sexualmente riesgosas, así como también la selección de los ítems de mayor utilidad para distinguir a dichas subpoblaciones (Holmes & Pierson, 2011), la identificación de diferentes clases asociadas al autoreporte de la adicción al tabaco (Muthen & Asparouhov, 2006), el estudio de las diferencias en cuanto a la interpretación de los ítems incluidos en los cuestionarios utilizados por los pacientes para reportar diferentes aspectos de su estado de salud (Sawatzky, Ratner, Kopec & Zumbo, 2012; Schmiede, Meek, Bryan & Petersen, 2012), la detección de examinados que dan autoreportes falsos en inventarios de personalidad (Holden & Book, 2009; Zickar, Gibby & Robie, 2004) y la validación de puntos de corte asociados a estándares de desempeño en pruebas estandarizadas (Jiao, Lissitz, Macready, Wang & Liang, 2011).

Así pues, el objetivo del presente artículo es ilustrar la utilidad de los modelos de mezcla de la Teoría de respuesta al ítem para detectar agrupaciones de examinados no observadas. Con ello, se pretende evidenciar la utilidad del modelo para la medición en el campo de las ciencias sociales y de la salud.

Método

Participantes

La muestra de examinados incluidos en el análisis corresponde a 2827 costarricenses cuyas edades oscilan entre los 60 y los 100 años, los cuales participaron en el estudio denominado Costa Rica Estudio de Longevidad y Envejecimiento Saludable (Rosero, Fernández & Dow, 2005):

In the first stage of the design model, a random selection was made from the database of the Census of Population of the 2000, totaling 9,600 individuals 55 years of age or older, after a stratification by five-year age groups that assures a sufficiently large number of observations for advanced ages. The sampling fraction in this selection varies between 1% for the ones born in 1941-1945 and 100% for the born ones before 1905. For the detailed longitudinal follow-up, including the survey to which the present

report refers, a sub-sampling was selected consisting of 60 "Areas of Health" (from a total of 102 in the whole country) aggregated into subregions. The sample covers 59% of the national territory (p. 2).

En relación con el consentimiento de los sujetos para participar en el estudio, Rosero, Fernández y Dow (2005, p. 4) afirman lo siguiente: "In the first visit, the participants granted their informed consent (Appendix 1) by means of their signature".

Instrumento

Como parte del estudio se aplicó una versión reducida de 15 ítems de la Escala de depresión geriátrica (Yesavage, Brink, Rose, Lum, Huang, Adey & Leirer, 1983), cuyas opciones de respuesta son 1 ("Sí") y 2 ("No"). La calidad técnica de dicho instrumento ha sido evaluada en una gran cantidad de estudios (Friedman, Heisel y Delavan, 2005; Brown & Schinka, 2005; Jongenelis, Pot, Eisses, Gerritsen, Derksen, Beekman & Ribbe, 2005) para medir el nivel de depresión en la población de adultos mayores.

Mediante este instrumento se midió el nivel de depresión de los examinados mediante preguntas como "¿Sintió que su vida está vacía?", "¿Estuvo preocupado o temiendo que algo malo le pasara?", entre otras. De este modo, quienes tienden a elegir la segunda opción de respuesta muestran menores niveles de depresión. Para interpretar las opciones de respuesta de la misma manera en todo el instrumento, se recodificaron los ítems 1, 5, 7, 11 y 13.

Procedimiento y análisis de datos

Los datos fueron descargados del sitio web mencionado en el apartado de participantes. Además, se eliminó de la muestra a todos los sujetos con datos faltantes en alguno de los 15 ítems, por lo que los análisis se realizaron con 1563 personas. Cabe señalar que para el objetivo del presente artículo, cual es el de ilustrar los posibles usos de un modelo psicométrico y no el de generalizar conclusiones hacia la población, no representa problema alguno la estrategia de eliminar examinados de la muestra.

En primer lugar se realizaron análisis descriptivos

para identificar la posible presencia de errores en la digitación de las respuestas. Posteriormente, se realizó un análisis factorial confirmatorio para recabar evidencias sobre la estructura interna de la escala. Finalmente, se analizaron los patrones de respuesta a los ítems mediante el modelo de mezcla de la Teoría de respuesta al ítem. Todos los análisis se ejecutaron con el programa MPLUS (Versión 7.3).

Resultados

En la tabla 1 se muestra para cada ítem el porcentaje de personas que eligió la categoría “No”, su correlación con la puntuación total (índice de discriminación), las cargas factoriales (con sus respectivos errores estándar) estimadas mediante un modelo unidimensional y la proporción de varianza explicada. La mayoría de personas eligió la opción “No”, además, tanto las cargas factoriales como los índices de ajuste son evidencias que apuntan a la unidimensionalidad del test. Cabe señalar que la confiabilidad de las puntuaciones derivadas de la aplicación del test estimada mediante el coeficiente alfa

de Cronbach es de .841.

Por otra parte, en la tabla 2 se pueden apreciar diferentes índices de ajuste para los tres modelos que se pusieron a prueba. Dado que actualmente no existe un único método ampliamente aceptado para comparar modelos con distintas clases latentes, es necesario combinar la teoría sustantiva que se utilizó como base para construir el test así como los diferentes índices de ajuste diseñados para comparar modelos con un número creciente de clases latentes (Masyn, 2013).

En este caso, se observa que el AIC, el BIC y el BIC ajustado (BICa) no coinciden en cuanto a cuál modelo representa mejor los datos, ya que el AIC más pequeño corresponde al modelo de tres clases, mientras que el BIC apunta al modelo de una clase y el BICa al de dos clases. Por su parte, las pruebas de razón de verosimilitudes Vuong-Lo-Mendell-Rubin (VLMR-LRT) y Lo-Mendell-Rubin ajustada (LMRA-LRT) muestran un valor p estadísticamente significativo ($p < .05$) cuando se compara el modelo de una clase con

Tabla 1

Descriptivos y resultados del análisis factorial confirmatorio (N = 1563)

Ítem	%	r	λ (EE)	R ²
1	0.897	.498	.788 (.026)	.622
2	0.784	.446	.629 (.030)	.396
3	0.736	.623	.859 (.017)	.738
4	0.786	.620	.851 (.018)	.724
5	0.883	.471	.772 (.026)	.596
6	0.807	.421	.610 (.031)	.372
7	0.883	.566	.864 (.020)	.746
8	0.828	.562	.799 (.022)	.638
9	0.547	.329	.475 (.033)	.225
10	0.788	.340	.494 (.035)	.244
11	0.956	.363	.721 (.041)	.520
12	0.864	.568	.810 (.023)	.655
13	0.867	.499	.743 (.027)	.552
14	0.865	.562	.812 (.023)	.659
15	0.561	.341	.493 (.031)	.243

Nota. $\chi^2 = 358.479$ (30, $p < .001$), CFI = .974, TLI = .97, RMSEA = .044 [.039, .048].

Tabla 2
Modelos TRI mixtos

Clases	Log-verosimilitud	Par	AIC	BIC	BICa	EN	VLMR-LRT	LMRA-LRT
1	-8790.878	30	17641.757	17802.388	17707.085	-	-	-
2	-8691.981	61	17505.961	17832.577	17638.794	.44	.0175	.0178
3	-8645.064	92	17474.128	17966.729	17674.466	.62	.1233	.1251

el de dos clases, no así con el de tres clases respecto del de dos clases (Asparouhov & Muthen, 2012). Finalmente, el valor de la entropía (EN) es moderado para el modelo de dos clases latentes, lo cual indica que existe cierta incertidumbre a la hora de asignar a los sujetos a una u otra clase latente (Asparouhov & Muthen, 2014).

Con base en estos resultados se decidió elegir el modelo de dos clases latentes para ejemplificar algunos de las etapas involucradas en este tipo de análisis. Sin embargo, es necesario recalcar que sería necesario en este punto incluir a expertos en el constructo medido por la escala para determinar desde un punto de vista sustantivo si interpretable la existencia de dos tipos de examinados en la muestra.

En la tabla 3 se presentan los parámetros estimados para cada ítem, tanto en la clase uno como en la clase dos. El parámetro *a* corresponde al índice de discriminación

y el parámetro *b*, al de dificultad. Entre paréntesis se consignan los errores estándar. En la última columna aparecen los valores de entropía para cada ítem, lo cual permite identificar qué tanto contribuye cada ítem a la separación entre las dos clases latentes.

Para complementar dicha información en la Figura 1 se presentan las probabilidades de elegir la opción “No” en cada ítem y para cada clase latente. En términos generales, se puede observar que las probabilidades de seleccionar dicha opción son menores en los examinados asignados a la clase 2, diferencia que se acentúa en los ítems 9 (“¿Prefirió quedarse en casa en vez de salir y hacer cosas?”) y 15 (“¿Creyó que las personas están en una situación mejor que usted?”), los cuales además muestran valores muy diferentes a los que se observan en los demás ítems en la tabla 1.

Conclusiones

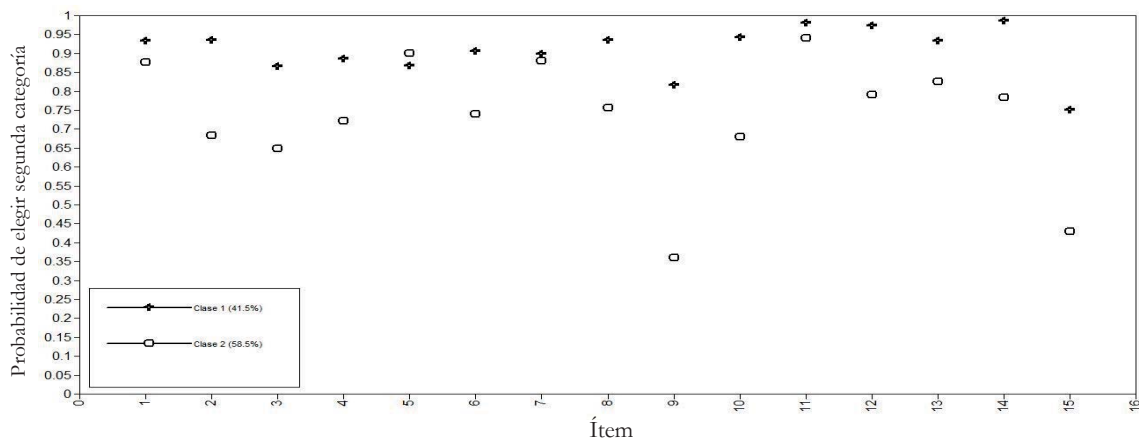


Figura 1. Probabilidades de elegir la segunda opción de respuesta.

Tabla 3
Parámetros estimados con el modelo de dos clases latentes

Ítem	Parámetro a		Parámetro b		Entropía univariada
	Clase 1	Clase 2	Clase 1	Clase 2	
1	2.962 (0.79)	2.219 (0.32)	-1.746 (0.176)	-1.480 (0.122)	.324
2	2.122 (0.59)	0.971 (0.17)	-1.968 (0.297)	-0.942 (0.150)	.324
3	1.564 (0.42)	3.543 (0.55)	-1.650 (0.405)	-0.430 (0.092)	.367
4	2.399 (0.49)	2.806 (0.33)	-1.496 (0.236)	-0.691 (0.085)	.355
5	2.692 (0.72)	3.526 (0.62)	-1.324 (0.152)	-1.440 (0.093)	.337
6	1.493 (0.39)	1.082 (0.17)	-2.030 (0.348)	-1.188 (0.138)	.315
7	4.038 (1.027)	3.869 (0.66)	-1.381 (0.104)	-1.296 (0.091)	.346
8	1.200 (0.491)	2.621 (0.37)	-2.689 (1.083)	-0.836 (0.099)	.340
9	1.209 (0.565)	0.364 (0.18)	-1.562 (0.375)	1.625 (1.184)	.329
10	1.011 (0.683)	0.625 (0.14)	-3.191 (1.458)	-1.314 (0.282)	.311
11	2.014 (0.613)	1.906 (0.28)	-2.776 (0.468)	-2.139 (0.188)	.302
12	3.826 (2.235)	1.886 (0.23)	-2.120 (0.174)	-1.101 (0.130)	.333
13	2.565 (0.711)	1.604 (0.23)	-1.820 (0.168)	-1.374 (0.117)	.322
14	3.959 (2.646)	2.070 (0.26)	-2.380 (0.272)	-1.025 (0.148)	.337
15	0.508 (0.211)	0.764 (0.13)	-2.297 (1.043)	0.419 (0.245)	.313

Nota. Entre paréntesis se presentan los errores estándar.

De acuerdo con los resultados observados, es posible plantear que los ítems 9 y 15 podrían estar midiendo de manera distinta el constructo en una escala sobre la cual se tienen evidencias de unidimensionalidad. A pesar de que la gran mayoría de los parámetros de los ítems no difieren de manera importante en las clases latentes identificadas, este análisis puede servir de base para utilizar un modelo confirmatorio (como por ejemplo, el análisis del funcionamiento diferencial del ítem) que permita entender mejor las razones por las cuales algunos ítems no miden de la misma manera el constructo en todos los examinados.

Un aspecto importante de recalcar es que los modelos de la Teoría de respuesta al ítem son de carácter exploratorio, por lo que los resultados derivados de su aplicación no deben emplearse como evidencia definitiva para la toma de decisiones. Es de vital importancia que se incluyan análisis descriptivos guiados por la teoría que fundamentó la construcción del instrumento para determinar características importantes de los sujetos asignados a cada clase latente. Y, finalmente, también se deben incorporar metodologías confirmatorias (utilizando otra muestra de examinados) para generar evidencias de que los resultados observados son generalizables.

Referencias

- AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. Washington, Estados Unidos: American Educational Research Association.
- Alexeev, N., Templin, J., & Cohen, A. (2011). Spurious Latent Classes in the Mixture Rasch Model. *Journal of Educational Measurement, 48*(3), 313-332.
- Asparouhov, T., & Muthen, B. (2012). *Using Mplus TECH11 and TECH14 to test the number of latent classes*. Recuperado de <https://www.statmodel.com/examples/webnotes/webnote14.pdf>
- Asparouhov, T., & Muthen, B. (2014). *Variable-Specific Entropy Contribution*. Recuperado de <http://www.statmodel.com/download/UnivariateEntropy.pdf>
- Baghaei, P., & Carstensen, C. (2013). Fitting the Mixed Rasch Model to a Reading Comprehension Test: Identifying Reader Types. *Practical Assessment, Research & Evaluation, 18*(5). Recuperado de <http://pareonline.net>
- Bolt, D., Cohen, A., & Wollackm J. (2001). A Mixture Item Response Model for Multiple-Choice Data. *Journal of Educational and Behavioral Statistics, 26*(4), 381-409.
- Bolt, D., Cohen, A., & Wollack, J. (2002). Item Parameter Estimation Under Conditions of Test Speededness: Application of a Mixture Rasch Model With Ordinal Constraints. *Journal of Educational Measurement, 39*(4), 331-348.
- Borsboom, D., Mellenbergh, G., & Heerden, J. (2004). The Concept of Validity. *Psychological Review, 111*(4), 1061-1071.
- Brown, L. M., & Schinka, J. A. (2005). Development and initial validation of a 15- item informant version of the Geriatric Depression Scale. *International journal of geriatric psychiatry, 20*(10), 911-918.
- Carter, N., Dalal, D., Lake, C., Lin, B., & Zickar, M. (2011). Using Mixed-Model Item Response Theory to Analyze Organizational Survey Responses: An Illustration Using the Job Descriptive Index. *Organizational Research Methods, 14*(1), 116 - 146.
- Chen, Y., & Jiao, H. (2013). Does Model Misspecification Lead to Spurious Latent Classes? An Evaluation of Model Comparison Indices. En R. Millsap, L. van der Ark, D. Bolt & C. Woods (Eds.), *New Developments in Quantitative Psychology. Presentations from the 77th Annual Psychometric Society Meeting* (pp. 345-355). Estados Unidos: Springer.
- Cho, S., & Cohen, A. (2010). A Multilevel Mixture IRT Model With an Application to DIF. *Journal of Educational and Behavioral Statistics, 35*(3), 336-370.
- Choi, Y., Alexeev, N., & Cohen, A. (2014). DIF Analysis using a Mixture 3PL Model with a Covariate on the TIMSS 2007 Mathematics Test. *KAERA Research Forum, 1*(1), 4-14. Recuperado de <http://www.kaera.org/research-forum/>
- Cohen, A., & Bolt, D. (2005). A Mixture Model

- Analysis of Differential Item Functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- Collins, L., & Lanza, S. (2010). *Latent Class and Latent Transition Analysis*. Estados Unidos: WILEY.
- Dai, Y. (2013). A Mixture Rasch Model With a Covariate: A Simulation Study via Bayesian Markov Chain Monte Carlo Estimation. *Applied Psychological Measurement*, 37(5), 375-396.
- De Ayala, R. (2009). *The Theory and Practice of Item Response Theory*. Estados Unidos: The Guilford Press.
- De Boeck, P., & Rijmen, F. (2003). A Latent Class Model for Individual Differences in the Interpretation of Conditionals. *Psychological Research*, 67, 219-231.
- DeMars, C., & Lau, A. (2011). Differential Item Functioning Detection with Latent Classes: How Accurately Can We Detect Who Is Responding Differentially? *Educational and Psychological Measurement*, 71(4), 597-616.
- Egberink, I., Meijer, R., & Veldkamp, B. (2010). Conscientiousness in the Workplace: Applying Mixture IRT to Investigate Scalability and Predictive Validity. *Journal of Research in Personality*, 44, 232- 244.
- Eid, M., & Rauber, M. (2000). Detecting Measurement Invariance in Organizational Surveys. *European Journal of Psychological Assessment*, 16(1), 20-30.
- Elosúa, P. & López, A. (2006). Clases latentes y funcionamiento diferencial del ítem. *Psicothema*, 17(3), 516-521.
- Embretson, S. (2007). Mixed Rasch Models for Measurement in Cognitive Psychology. En M. von Davier & C. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications* (pp. 235-253). Estados Unidos: Springer.
- Formann, A., & Kohlmann, T. (2002). Three-Parameter Linear Logistic Latent Class Analysis. En J. Hagenars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 183 - 210). Estados Unidos: Cambridge University Press.
- Frick, H., Strobl, C., & Zeileis, A. (en prensa). Rasch Mixture Models for DIF Detection: A Comparison of Old and New Score Specifications. *Educational and Psychological Measurement*. Recuperado de <http://epm.sagepub.com>
- Friedman, B., Heisel, M. J., & Delavan, R. L. (2005). Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatrics Society*, 53(9), 1570-1576.
- Gollwitzer, M., Eid, M., & Jürgensen, R. (2005). Response Styles in the Assessment of Anger Expression. *Psychological Assessment*, 17(1), 56-59.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Estados Unidos: Sage.
- Holden, R., & Book, A. (2009). Using Hybrid Rasch-Latent Class Modeling to Improve the Detection of Fakers on a Personality Inventory. *Personality and Individual Differences*, 47, 185-190.
- Holmes, W., & Pierson, E. (2011). A Mixture IRT Analysis of Risky Youth Behavior. *Frontiers in Psychology*, 2, 1-10.
- Hong, S., & Min, S. (2007). Mixed Rasch Modeling of the Self-Rating Depression Scale: Incorporating Latent Class and Rasch Rating Scale Models. *Educational and Psychological Measurement*, 67(2), 280-299.
- Jiao, H., Lissitz, R., Macready, G., Wang, S., & Liang, S. (2011). Exploring Levels of Performance Using the Mixture Rasch Model for Standard Setting. *Psychological Test and Assessment Modeling*, 53(4), 499-522.
- Jongenelis, K., Pot, A. M., Eisses, A. M. H., Gerritsen, D. L., Derksen, M., Beekman, A. T. F & Ribbe, M. W. (2005). Diagnostic accuracy of the original 30-item and shortened versions of the Geriatric Depression Scale in nursing home patients. *International journal of geriatric psychiatry*, 20(11), 1067-1074.
- Lee, H. & Beretvas, S. (2014). Evaluation of Two Types of Differential Item Functioning in Factor Mixture

- Models With Binary Outcomes. *Educational and Psychological Measurement*, 74(5), 831-858.
- Li, F., Cohen, A., Kim, S., & Cho, S. (2009). Model Selection Methods for Mixture Dichotomous IRT Models. *Applied Psychological Measurement*, 33(5), 353-373.
- Lubke, G., & Muthen, B. (2005). Investigating Population Heterogeneity with Factor Mixture Models. *Psychological Methods*, 10(1), 21-39.
- Masyn, K. (2013). Latent Class Analysis and Finite Mixture Modeling. En T. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology: Volumen 2* (pp. 551 - 611). Estados Unidos: Oxford University Press.
- Meij, A., Kelderman, H., & Flier, H. (2008). Fitting a Mixture Item Response Theory Model to Personality Questionnaire Data: Characterizing Latent Classes and Investigating Possibilities for Improving Prediction. *Applied Psychological Measurement*, 32(8), 611-631.
- Meij, A., Kelderman, H., & Flier, H. (2010). Improvement in Detection of Differential Item Functioning Using a Mixture Item Response Theory Model. *Multivariate Behavioral Research*, 45, 975 - 999.
- Meiser, T., & Machunsky, M. (2008). The Personal Structure of Personal Need for Structure. *European Journal of Psychological Assessment*, 24(1), 27-34.
- Meyer, J. (2010). A Mixture Rasch Model with Item Response Time Components. *Applied Psychological Measurement*, 34(7), 521-538.
- Mislevy, R., & Verhelst, N. (1990). Modeling Item Responses When Different Subjects Employ Different Solution Strategies. *Psychometrika*, 55(2), 195-215.
- Mittelhaeuser, M., Béguin, A., & Sijtsma, K. (2013). Modeling Differences in Test-Taking Motivation: Exploring the Usefulness of the Mixture Rasch Model and Person-Fit Statistics. En R. Millsap, L. van der Ark, D. Bolt & C. Woods (Eds.), *New Developments in Quantitative Psychology. Presentations from the 77th Annual Psychometric Society Meeting* (pp. 345-355). Estados Unidos: Springer.
- Mneimneh, Z., Heeringa, S., Tourangeau, R., & Elliott, M. (2014). Bringing Psychometrics and Survey Methodology: Can Mixed Rasch Models Identify Socially Desirable Reporting Behavior? *Journal of Survey Statistics and Methodology*, 2, 257-282.
- MPLUS (Versión 7.3) [Software de computación]. Los Ángeles, CA: Muthen & Muthen.
- Muthen, B., & Asparouhov, T. (2006). Item Response Mixture Modeling: Application to Tobacco Dependence Criteria. *Addictive Behaviors*, 31, 1050-1066.
- Muthen, B., & Lubke, G. (2007). Performance of Factor Mixture Models as a Function of Model Size, Covariate Effects, and Class-Specific Parameters. *Structural Equation Modeling*, 14(1), 26-47.
- Nylund, K., Asparouhov, T., & Muthen, B. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 14(4), 535-569.
- Oliveri, M., Ercikan, K., Zumbo, B., & Lawless, R. (2014). Uncovering Substantive Patterns in Student Responses in International Large-Scale Assessments? Comparing a Latent Class to a Manifest DIF Approach. *International Journal of Testing*, 14, 265-287.
- Paek, I., & Cho, S. (2015). A Note on Parameter Estimate Comparability: Across Latent Classes in Mixture IRT Modeling. *Applied Psychological Measurement*, 39(2), 135-143.
- Preinerstorfer, D., & Formann, A. (2011). Parameter Recovery and Model Selection in Mixed Rasch Models. *British Journal of Mathematical and Statistical Psychology*, 65, 251-262.
- Raykov, T., & Marcoulides, G. (2011). *Introduction to Psychometric Theory*. Estados Unidos: Routledge.
- Raykov, T., Marcoulides, G., Lee, C., & Chang, C. (2013). Studying Differential Item Functioning via

- Latent Variable Modeling: A Note on a Multiple-Testing Procedure. *Educational and Psychological Measurement*, 73(5), 898-908.
- Rosero, L., Fernández, X., & Dow, W. (2005). CRELES: Costa Rican Longevity and Healthy Aging Study (Costa Rica Estudio de Longevidad y Envejecimiento Saludable). ICPSR26681-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distribuidor]. Recuperado de <http://doi.org/10.3886/ICPSR26681.v2>
- Rost, J. (1990). Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- Sawatzky, R., Ratner, P., Kopec, J., & Zumbo, B. (2012). Latent Variable Mixture Models: A Promising Approach for the Validation of Patient Reported Outcomes. *Quality of Life Research*, 21, 637-650.
- Schmiege, S., Meek, P., Bryan, A., & Petersen, H. (2012). Latent Variable Mixture Modeling: A Flexible Statistical Approach for Identifying and Classifying Heterogeneity. *Nursing Research*, 61(3), 204-212.
- Sterba, S. (2013). Understanding Linkages Among Mixture Models. *Multivariate Behavioral Research*, 48, 775-815.
- Smit, A., Kelderman, H., & van der Flier, H. (1999). Collateral Information and Mixed Rasch Models. *Methods of Psychological Research*, 4(3). Recuperado de <http://dare.ubvu.vu.nl/handle/1871/18667>
- Smit, A., Kelderman, H., & van der Flier, H. (2000). The Mixed Birnbaum Model: Estimation Using Collateral Information. *Methods of Psychological Research*, 5(4). Recuperado de <http://dare.ubvu.vu.nl/handle/1871/18670>
- Tay, L., Newman, D., & Vermunt, J. (2011). Using Mixed-Measurement Item Response Theory With Covariates (MM-IRT-C) to Ascertain Observed and Unobserved Measurement Equivalence. *Organizational Research Methods*, 14(1), 147-176.
- Tueller, S., & Lubke, G. (2010). Evaluation of Structural Equation Mixture Models: Parameter Estimates and Correct Class Assignment. *Structural Equation Modeling*, 17, 165-192.
- Vermunt, J. (2014). Latent Class Model. En A. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3509 - 3515). Estados Unidos: Springer.
- Von Davier, M., & Molenaar, I. (2003). A Person-Fit Index for Polytomous Rasch Models, Latent Class Models, and their Mixture Generalizations. *Psychometrika*, 68(2), 213-228.
- Von Davier, M., & Yamamoto, K. (2007). Mixture-Distribution and HYBRID Rasch Models. En M. von Davier & C. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications* (pp. 99-115). Estados Unidos: Springer.
- Von Davier, M. (2010). Mixture Distribution Item Response Theory, Latent Class Analysis, and Diagnostic Mixture Models. En S. Embretson (Ed.), *Measuring Psychological Constructs: Advances in Model-Based Approaches* (pp. 11-34). Estados Unidos: American Psychological Association.
- Willse, J. (2011). Mixture Rasch Models with Joint Maximum Likelihood Estimation. *Educational and Psychological Measurement*, 71(1), 5-19.
- Zickar, M., Gibby, R., & Robie, C. (2004). Uncovering Faking Samples in Applicant, Incumbent, and Experimental Data Sets: An Application of Mixed-Model Item Response Theory. *Organizational Research Methods*, 7(2), 168-190.

Recibido: 7 de abril de 2015

Aceptado: 16 de setiembre de 2015