

ARTÍCULO DE INVESTIGACIÓN

LA MEDIDA DE LA PERCEPCIÓN EMOCIONAL CON EL MODELO DE RASCH

MEASURING EMOTION PERCEPTION WITH THE RASCH MODEL

ANA R. DELGADO*

UNIVERSIDAD DE SALAMANCA, SALAMANCA-ESPAÑA

FECHA RECEPCIÓN: 20/3/2016 • FECHA ACEPTACIÓN: 4/5/2016

Para citar este artículo: Delgado, A. R. (2016). La medida de la percepción emocional con el modelo de Rasch. *Psychologia: Avances de la Disciplina*, 10(2), 27-32.

Resumen

Reconocer y denominar expresiones emocionales son habilidades en cuyo estudio existe un creciente interés por parte de la psicología y las neurociencias; trastornos tales como el autismo, la esquizofrenia y la depresión se caracterizan por dificultades en reconocer las emociones ajenas. El hallazgo de diferencias a favor de las mujeres en Percepción Emocional (PE) ha sido replicado transculturalmente por lo que sirve como evidencia de validez. En esta investigación, de diseño instrumental, se construyeron 28 ítems y se pusieron a prueba mediante el modelo de Rasch (MR) en una muestra de población general española (102 varones y 102 mujeres; rango de edad=18-65). El ajuste datos-modelo fue suficiente. Las medidas de las personas muestran que el test resultó fácil ($range = -.60$ a 4.14 logits). Se hallaron las diferencias sexuales predichas, $d = -.36$ (d con corrección de la atenuación = $-.50$); que ningún ítem mostrase funcionamiento diferencial (DIF) asociado al sexo y que la variable construida tenga propiedades intervalares añada confianza al valor estimado del tamaño del efecto. La construcción de una escala de intervalo en la que se localizan conjuntamente ítems y personas ofrece información cuantitativa y permite detectar patrones inesperados de respuesta que pueden ser de utilidad para el diagnóstico neuropsicológico.

Palabras clave: Inteligencia emocional; ítems de elección múltiple; modelo de Rasch percepción emocional; diferencias sexuales.

Abstract

Recognizing and labeling emotion expressions are abilities in whose study psychology and the neurosciences are increasingly interested; disorders such as autism, schizophrenia and depression are characterized by difficulties in recognizing others' emotions. The finding of sex differences favoring women in Emotion Perception (EP) has been replicated cross-culturally, and so it can be used as validity evidence. In this instrumental design study, some 28 items were constructed and tested by means of the Rasch model (RM) in a Spanish community sample (102males and 102 females; age range=18-65). Data-model fit was good enough. Person measures show that the test was easy ($range = -.60$ to 4.14 logits). Even so, the predicted sex differences favoring women were found, $d = -.36$

* Agradezco al Centro Documental de la Memoria Histórica (CDMH, Salamanca, España) que me permitiera utilizar un espacio para aplicar pruebas en su museo; y a Ron Dotsch, Gijs Bijlstra y Oliver Langner el permiso para emplear la Radboud Faces Database. La correspondencia puede enviarse a Ana R. Delgado. Facultad de Psicología. Avenida de la Merced 109-131. 37005 Salamanca (Spain). E-mail: adelgado@usal.es.

(disattenuated $d = -.50$). That no item showed sex-related Differential Item Functioning (DIF) and that the constructed variable has interval scale properties add confidence to the effect size estimated value. The construction of an interval scale on which EP items and persons are conjointly represented offers quantitative information and allows the discovery of unexpected response patterns.

Keywords: Emotional intelligence; emotion perception; multiple-choice items; Rasch model; sex differences.

Introducción

Reconocer y denominar expresiones emocionales son habilidades en cuyo estudio existe un creciente interés por parte de la psicología y las neurociencias (Delgado, 2012; Marquez & Delgado, 2012; Mathias & Burke, 2009). El adecuado reconocimiento de expresiones de emoción en otros es clave en la interacción social; existe evidencia de que trastornos tales como el autismo, la esquizofrenia y la depresión se caracterizan por dificultades en reconocer las emociones ajenas (Griffiths & Ashwin, 2016).

Las tareas de reconocimiento emocional habitualmente empleadas suelen incluir ítems de siete «familias» de emociones discretas cuyas etiquetas verbales son alegría, pena, sorpresa, miedo, asco, rabia y desprecio (Delgado, 2009; Marquez & Delgado, 2012; Merten, 2005; Tracy & Randles, 2011) o términos similares. En los ítems de elección múltiple, se incluyen, además, imágenes de expresiones neutras. Puesto que en el contexto de la investigación de la inteligencia emocional se habla de Percepción Emocional (PE) para describir tareas como la anteriormente descrita (Mayer, Roberts, & Barsade, 2008), ambas denominaciones se emplearán de forma intercambiable a lo largo de este artículo.

Pese a que las tareas de PE son, en general, excesivamente fáciles (Thompson & Voyer, 2014) y poco fiables (Gignac, 2009), un hallazgo robusto y replicado transculturalmente es que existen diferencias en PE favorables a las mujeres (Hall, Gunnery, & Horgan, 2016; Merten, 2005; Thompson & Voyer, 2014). Tres revisiones no solapadas han mostrado evidencia indiscutible de la superioridad femenina en la precisión de los juicios afectivos (Hall, Gunnery, & Horgan, 2016). Un metaanálisis más sofisticado estadísticamente que sus predecesores estima que el tamaño del efecto «verdadero»

de la diferencia en PE sería $d = .23$ (Thompson & Voyer, 2014).

La interpretación cuantitativa de esa diferencia asume que las puntuaciones obtenidas de las tareas de PE se encuentran, al menos, en escala intervalar, pero no se aporta evidencia del cumplimiento de ese supuesto que la aproximación basada en el modelo de Rasch sí permite poner a prueba. Por ello, el objetivo de este trabajo fue la construcción y validación inicial de una prueba de PE desde la aproximación Rasch, que cuenta con notables ventajas sobre la teoría clásica de los tests, e. g., si hay suficiente ajuste datos-modelo, entonces la medida de las personas no depende de los ítems empleados, las estimaciones de dificultad de los ítems son independientes de las personas medidas, y es posible construir una escala intervalar conjunta y llevar a cabo análisis paramétricos (Engelhard, 2013; Gómez-Rada, 2010; Prieto & Delgado, 2003; Rasch, 1960).

Método

Participantes

La muestra se compuso de doscientos cuatro voluntarios de lengua materna española, de edades comprendidas entre los 18 y los 65 años, e igual número de varones y mujeres. Se contó con una muestra heterogénea procedente de la población general, aunque el modelo de Rasch no exige muestras representativas (i. e., los valores de localización de los ítems son independientes de las personas medidas).

Procedimiento

El test se aplicó individualmente en un ordenador portátil, sin limitaciones de tiempo y siguiendo los requisitos éticos.

Instrumentos

Se construyó un test informatizado con 28 ítems, 4 para cada una de las 7 emociones discretas: alegría, pena, sorpresa, miedo, asco, rabia y desprecio. La mitad de los ítems para cada emoción constaban de una etiqueta (e. g., sorpresa) de entre las siete mencionadas y ocho opciones de respuesta consistentes en fotografías de la cara de un modelo (la mitad de las veces masculino, la otra mitad, femenino) posando con las siete expresiones emocionales además de la neutral. La otra mitad de los ítems constaban de la cara de un modelo (la mitad de las veces masculino, la otra mitad, femenino) posando con una de las expresiones emocionales y las ocho opciones de respuesta verbales: alegría, pena, sorpresa, miedo, asco, rabia, desprecio, neutral. Las fotografías se seleccionaron de la Radboud Faces Database (Langner et al., 2010). El posible efecto del sexo del modelo y del tipo de ítem sobre la dificultad se controló por inclusión, aunque resultados como los del metaanálisis de Thompson y Voyer (2014) permiten predecir que variaciones de ese tipo no darán lugar a diferencias en la dificultad de los ítems.

El test se programó en LiveCode 4.6 (2011). Los datos de identificación, sexo y edad del participante, así como la opción de respuesta elegida para cada ítem y el acierto/error se almacenan automáticamente en matrices listas para exportar.

Análisis de datos

Para este estudio, de diseño instrumental, se empleó el modelo de Rasch (MR), que pertenece a la tradición de los modelos de medición invariante (Engelhard, 2013). En el contexto de la medida de una aptitud, la probabilidad de que el sujeto n acierte el ítem dicotómico i es, según el MR:

$$P_{ni} = \exp(B_n - D_i) / (1 + \exp(B_n - D_i))$$

B_n es el nivel del sujeto n

D_i es la dificultad del ítem i

Los análisis se llevaron a cabo con el programa Winsteps 3.80.1 (Linacre, 2013). El ajuste datos-

modelo se valora con los indicadores *infit* y *outfit* calculados a partir de los residuos. Según las indicaciones de Winsteps 3.80.1 (Linacre, 2013), valores de *infit* / *outfit* entre 0.5 y 1.5 son adecuados; valores entre 1.5 y 2.0 son improductivos, pero no degradan la medición y valores superiores a 2.0 degradarían las medidas. Los análisis del funcionamiento diferencial de los ítems (DIF, siguiendo el acrónimo en inglés) se llevan a cabo para contrastar la validez generalizada de las medidas para distintos grupos, en este caso, para varones y mujeres: las calibraciones de los ítems en ambos grupos se contrastan con el estadístico t de Welch usando niveles *alfa* de confianza con la corrección de Bonferroni (Linacre, 2013).

Resultados

La menor proporción de aciertos se dio en el caso del desprecio (.71), seguida del miedo (.84), la pena (.92) y el asco (.92), la rabia (.95), la alegría (.98) y la sorpresa (.99). El estimador de la consistencia interna *alfa* de Cronbach fue .75.

El análisis Rasch no mostró ninguna correlación negativa ítem-medida. El ítem 15 (de alegría) fue acertado por todos, por lo que su dificultad no puede ser estimada; el 19 % de las personas acertaron los 28 ítems.

El ajuste resultó suficiente como para proseguir el análisis. Para los ítems, el valor promedio de *infit* fue .99 ($DT = .09$) y el de *outfit* 1.08 ($DT = .50$). Para las personas, el valor promedio de *infit* fue 1.00 ($DT = .29$) y el de *outfit* 1.08 ($DT = 1.16$). Ningún ítem o persona mostró *infit* superior a 2. Dos ítems (de sorpresa) mostraron *outfit* superior a 2, pero los tests estadísticos asociados no alcanzaron significación.

En cuanto a las puntuaciones de los sujetos, el 13% tuvieron *outfit* > 2. Para los sujetos con puntuaciones imperfectas, el nivel promedio en unidades *logit* fue 2.73; *rango* = $-.60$ a 4.14. Los indicadores globales de fiabilidad del modelo fueron .93 para los ítems y .51 para las personas. La tabla 1 muestra las características de los ítems y los resultados principales del análisis Rasch.

Tabla 1. Resultados del análisis Rasch de los ítems

Ítem	Emoción	Modelo	Rasch D _i	ET	Infit	Outfit
1	alegría	femenino	-1.32	.47	1.00	1.74
2	sorpresa	masculino	.50	.24	1.07	1.01
3	pena	femenino	1.62	.19	1.02	1.05
4	alegría	masculino	-1.56	.52	.94	1.03
5	sorpresa	femenino	-.95	.40	.98	1.37
6	rabia	masculino	-.05	.29	1.14	1.96
7	asco	femenino	-.66	.36	.99	.99
8	miedo	masculino	.11	.27	.90	1.23
9	desprecio	femenino	2.07	.18	1.02	1.12
10	pena	masculino	-.43	.33	.96	.67
11	rabia	femenino	-.66	.36	1.01	.64
12	asco	masculino	-.66	.36	.99	1.61
13	miedo	femenino	-1.32	.47	.90	.47
14	desprecio	masculino	2.07	.18	.83	.77
15	alegría	femenino	---	---	---	---
16	sorpresa	masculino	-1.12	.43	1.01	2.19
17	pena	femenino	-1.56	.52	1.00	.75
18	alegría	masculino	-1.56	.52	.91	.40
19	sorpresa	femenino	-.80	.38	1.03	2.15
20	rabia	masculino	-1.32	.47	1.01	.41
21	asco	femenino	-.14	.30	1.15	1.09
22	miedo	masculino	3.02	.18	1.08	1.11
23	desprecio	femenino	1.70	.19	.81	.69
24	pena	masculino	-.95	.40	.97	.58
25	rabia	femenino	.11	.27	.91	.64
26	asco	masculino	1.28	.20	1.22	1.43
27	miedo	femenino	.38	.25	1.07	1.28
28	desprecio	masculino	2.20	.18	.83	.73

No se han hallado efectos estadísticamente significativos, sobre la dificultad de los ítems, ni del sexo del modelo fotográfico, $t(25) = -.43, p = .67$, ni del tipo de ítem, $t(25) = .35, p = .73$. El mapa de la variable, o mapa Wright, puede verse en la tabla 2.

Tabla 2. Wright map.

Medida	Persona - MAPA - Ítem
5	.##### T+
4	.##### +
3	.##### S
2	.##### M T
1	.##### S
0	.##### S
-1	.##### S
-2	.##### S

cada "#" es 4: cada "." es 1 a 3

No se halló DIF asociado al sexo y por tanto las diferencias sexuales en PE pueden contrastarse con confianza: $t(202) = 2.57, p = .01$. Las mujeres fueron superiores a los varones, $M(f) = 3.50, DT = 1.34; M(m) = 2.98, DT = 1.54, d = -.36$. Empleando la fórmula de corrección de la atenuación (dividiendo el tamaño del efecto por la raíz cuadrada del estimador de fiabilidad; Hunter and Schmidt, 1990) se obtiene $d = -.50$, un tamaño medio.

Discusión

La emoción que ha dado lugar a los ítems más difíciles, a juzgar por la proporción de aciertos, es el desprecio, seguida del miedo, la pena y el asco, la rabia, la alegría y la sorpresa. El estimador de la consistencia interna *alfa* de Cronbach fue .75, valor que, de acuerdo con los estándares clásicos, resultaría adecuado para las puntuaciones de un test, siempre que no fuera a ser empleado para decisiones de muy alto nivel.

El ajuste de los datos al MR resultó suficiente, aunque se detectaron patrones inesperados en las respuestas de algunos sujetos, lo que en algunos casos puede dar lugar a hipótesis clínicas sobre las posibles razones del desajuste. Las puntuaciones perfectas (i. e., las de sujetos que aciertan todos los ítems) no ofrecen información para la estimación de los parámetros. Para los sujetos con puntuaciones imperfectas, el nivel promedio en unidades *logit* fue alto, i. e., se trata de un test fácil para la población general, aunque con un rango amplio. En cuanto a los indicadores globales de fiabilidad del modelo fueron .93 para los ítems y .51 para las personas, un valor bajo, pero habitual en este tipo de tarea (Gignac, 2009).

La dificultad de los ítems no se vio afectada ni por el sexo del modelo fotográfico, ni por el tipo de ítem. Tampoco se halló DIF asociado al sexo, lo que indica validez generalizada de los ítems para ambos sexos y permite contrastar estadísticamente las diferencias sexuales en PE sin que la calidad técnica de los ítems pueda servir como hipótesis alternativa plausible.

Como ya se ha señalado en la introducción, existen diferencias por sexo en EP en varias culturas (Merten, 2005) y el hecho de que se hayan replicado en este estudio es un indicador de validez. Que las puntuaciones Rasch tengan propiedades intervalares permite hacer una interpretación cuantitativa de la diferencia. Las mujeres fueron superiores a los varones en EP, como se esperaba a partir de los robustos resultados de estudios previos (Hall, Gunnery, & Horgan, 2016; Merten, 2005; Thompson & Voyer, 2014). Dada la baja fiabilidad de las puntuaciones, habitual en este tipo de tarea que tan fácil resulta en la población general, es recomendable emplear la fórmula de corrección de la atenuación para estimar el tamaño del efecto «real», i. e., el valor que se alcanzaría si se contara con un instrumento que midiera sin error

(Gignac, 2009) que es, en este caso, de tamaño medio. No obstante, sería deseable contar con ítems más difíciles en futuras versiones.

Finalmente, una ventaja añadida de los modelos de tipo Rasch, que resulta de especial interés para el diagnóstico neuropsicológico, es que permiten detectar patrones inesperados de respuesta para individuos concretos, e. g., una persona que obtiene una alta puntuación, pero falla todos los ítems de la «familia» asco, aparecería como desajustada, lo que permite plantear hipótesis de interés clínico (Prieto, Delgado, Perea, & Ladera, 2010).

Referencias

- Delgado, A.R. (2009). Spanish Basic Emotion Words are Consistently Ordered. *Quality & Quantity*, 43, 509-517. doi:10.1007/s11135-007-9121-3
- Delgado, A. (2012). Using the Rasch Model to Test the Psychometric Quality of Emotion Recognition Items. *Journal of the International Neuropsychological Society*, 18 (S1), 69.
- Engelhard, G. (2013). *Invariant Measurement: Using Rasch Models in the Social, Behavioral and Health Sciences*. New York, NY: Routledge.
- Gignac, G. E. (2009). Psychometrics and the Measurement of Emotional Intelligence. In C. Stough, D. Saklofske, & J. Parker (Ed.), *Assessing Emotional Intelligence: Theory, Research & Applications* (pp. 9-41). New York: Springer.
- Gómez-Rada, C.A. (2010). Diseño, construcción y validación de un instrumento que evalúa la calidad de vida laboral percibida en organizaciones colombianas, desde la teoría de respuesta al ítem. *Psychologia: Avances de la Disciplina*, 4(11), 13-124.
- Griffiths, P., & Ashwin, C. (2016). Accuracy in Perceiving Facial Expressions of Emotion in Psychopathology. In J. A. Hall, M. S. Mast and T. V. West (eds.), *The Social Psychology of Perceiving Others Accurately* (pp. 185-205). Cambridge: Cambridge University Press.
- Hall, J. A., Gunnery, S. D., & Horgan, T. G. (2016). Gender Differences in Interpersonal Accuracy. In J. A. Hall, M. S. Mast and T. V. West (eds.), *The Social*

- Psychology of Perceiving Others Accurately* (pp. 309-327). Cambridge: Cambridge University Press.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage Publications.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and Validation of the Radboud Faces Database. *Cognition and Emotion*, 24, (8), 1377-1388. doi:10.1080/02699930903485076
- Linacre, J. M. (2013). *Winsteps Rasch Measurement Computer Program*, version 3.80.1. Chicago: Winsteps.com.
- LiveCode 4.6 [computer software]. (2011). Retrieved from <http://runrev.com/>
- Marquez, M. G., & Delgado, A. R. (2012). Revisión de las medidas de reconocimiento y producción de emociones. *Anales de Psicología*, 28(3), 978-985. doi:10.6018/analesps.28.3.148961
- Mathias, J. L., & Burke, J. (2009). Cognitive Functioning in Alzheimer's and Vascular Dementia: A Meta-Analysis. *Neuropsychology*, 23(4), 411-423. doi: 10.1037/a0015384
- Mayer, J. D., Roberts, R. D. & Barsade, S. G. (2008). Human Abilities: Emotional Intelligence. *Annual Review of Psychology*, 59, 507-536. doi: 10.1146/annurev.psych.59.103006.093646
- Merten, J. (2005). Culture, Gender and the Recognition of the Basic Emotions. *Psychologia*, 48, 306-316. doi: 10.2117/psysoc.2005.306
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Prieto, G., Delgado, A. R., Perea M. V., & Ladera, V. (2010). Scoring Neuropsychological Tests Using the Rasch Model: An Illustrative Example With the Rey-Osterreith Complex Figure. *The Clinical Neuropsychologist*, 24(1), 45-56. doi: 10.1080/13854040903074645
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Thompson, A. E., & Voyer, D. (2014). Sex Differences in the Ability to Recognise Non-Verbal Displays of Emotion: a Meta-Analysis. *Cognition and Emotion*, 28(7), 1164-1195. doi: 10.1080/02699931.2013.875889.