



MODELOS POLITÓMICOS DE LA TEORÍA DE RESPUESTA AL ÍTEM APLICADOS A LA EVALUACIÓN DE LA ACÚSTICA DE SALAS MEDIANTE ENCUESTAS.

Salvador Cerdá*, Carlos Abellán, Alicia Giménez¹
Jaume Segura, Rosa M. Cibrián²

¹ Universitat Politècnica de València

² Universitat de València

RESUMEN

Los modelos politómicos son herramientas utilizadas en la teoría de respuesta al ítem (TRI) para medir habilidades o características de los individuos a través de pruebas. A diferencia de los modelos dicotómicos, los modelos politómicos permiten respuestas con múltiples categorías. En este trabajo estudiamos la utilidad de este tipo de análisis para analizar los resultados de las encuestas de valoración de salas desde el punto de vista acústico. Estudiamos los resultados que se obtienen usando las principales librerías desarrolladas en el entorno estadístico *RStudio* y proponemos el uso de una de ellas en particular.

ABSTRACT

Polytomous models are tools used in item response theory (IRT) to measure abilities or characteristics of individuals through tests. Unlike dichotomous models that only consider correct or incorrect responses, polytomous models allow responses with multiple categories. In this work we study the usefulness of this type of analysis to study the results of room assessment surveys from an acoustic point of view. We analyze the results obtained using the main libraries developed in the *RStudio* statistical environment.

Palabras Clave— teoría de respuesta al ítem, evaluación subjetiva de salas, selección de modelos

1. INTRODUCCIÓN

La teoría de la respuesta al ítem (IRT) o la teoría del rasgo latente (LTT) es una teoría psicométrica,

utilizada para el análisis de pruebas psicológicas y pruebas generales. Estudia la relación entre un conjunto de datos obtenidos en un proceso de medición (respuestas a ítems de una prueba) con determinadas dimensiones latentes, como los rasgos de personalidad de los sujetos a los que se les ha administrado. En este sentido, proporciona un modelo matemático que presenta de forma simplificada e idealizada el comportamiento de los sujetos y las preguntas. En el campo de las pruebas de habilidad, tenemos un conjunto de preguntas que son formuladas por un grupo de personas. Las preguntas pueden ser correctas o incorrectas. Según los resultados podemos establecer, por un lado, la dificultad de las preguntas, observando el porcentaje de aciertos. Por otro lado, podemos establecer la capacidad de las personas al ver los éxitos que tienen. Con ligeras modificaciones, se pueden considerar preguntas con una calificación de respuesta categórica. La TRI busca encontrar una variable continua que modele la dificultad de las preguntas y las habilidades de las personas. El modelo ajusta esta nueva variable (latente) a cada individuo de tal manera que las probabilidades de responder a cada categoría están relacionadas con el valor de la dimensión latente. En tal caso, se dice que la dimensión latente separa a los individuos de las preguntas. Si las preguntas admiten sólo dos categorías posibles, el modelo utilizado se denomina dicotómico, por ejemplo, Rasch (1960) considera [1]:

$$P(Y_{p,i} = 1 | \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

en el que la probabilidad de responder correctamente a una pregunta se determina en función de la dimensión o habilidad latente de la persona θ_p y la dificultad de la pregunta β_i . En

nuestro caso donde tenemos más de dos categorías en las respuestas, usamos el modelo de crédito parcial (PCM), introducido en 1982 por Masters [2], quien descompuso la respuesta a un ítem en una serie de pares ordenados de categorías adyacentes, luego aplicó un modelo dicotómico a cada par asumiendo discriminaciones iguales entre los ítems. En otras palabras, tenemos que trabajar con un modelo politómico y usamos el modelo de crédito parcial [3]:

$$P(Y_{p,i} = x) = \frac{\exp \sum_{k=0}^x (\theta_p - \tau_{ki})}{\sum_{j=0}^{m_i} \exp \sum_{k=0}^j (\theta_p - \tau_{ki})}$$

siendo: m_i el número de posibles respuestas, x la respuesta al ítem, i el índice del ítem, θ_p la habilidad de la persona y τ_{ki} el parámetro límite para el paso k en el ítem i .

Este modelo está implementado en R en diferentes librerías. En este trabajo valoramos las diferentes librerías analizando los resultados y estableciendo la conveniencia del uso de alguna de ellas. Para ello se utiliza como criterio la correlación entre la variable latente y las puntuaciones experimentales obtenidas en los ítems. Hemos comparado las librerías “cdm” [4], “eRm” [7], “ltm” [8], “mirt” [3] y “TAM” [10], “psychotools” [11] y “sirt” [9]. Para determinar las habilidades de cada individuo se ha optado por la estimación esperada a posteriori (EAP): La estimación EAP es un enfoque bayesiano que combina la distribución previa de las habilidades de la persona con la probabilidad de observar las respuestas. Proporciona un promedio ponderado de los valores de capacidad más probables en función de las respuestas de la persona y la distribución de capacidades en la población [2].

Para la valoración subjetiva de la acústica de salas se ha utilizado una encuesta de nuestro diseño, que se pasó a los asistentes a diferentes conciertos celebrados en 8 teatros, salas de conciertos y auditorios de la Comunidad Valenciana ubicados en las localidades: Valencia, Alicante, Ribarroja, Benaguacil, Liria, Torrent, Algemesí y Castelló. La encuesta compuesta por 58 preguntas se divide en 6 apartados, 3 de ellos directamente relacionados con la sala y la percepción acústica en ella y otros 3 correspondientes a datos sociológicos, preferencias musicales, etc. del oyente y, por tanto, directamente relacionados con

la persona que responde la encuesta. Los tres apartados relacionados con la sala y su acústica, que engloban 48 preguntas, y que son los que analizaremos en este trabajo, corresponden a: Apartado A, “Aspectos generales”; Apartado B, “Percepción acústica detallada”, y Apartado C, “Percepción acústica global”. La encuesta ha sido respondida por 704 personas, de las cuales 184 corresponden a expertos en música, no en acústica de salas, y 520 al público en general asistente al concierto [5], [6].

2. RESULTADOS

Para el análisis de las encuestas hemos considerado que para cada una de las secciones de nuestra encuesta vamos a calcular la variable latente correspondiente. Esta variable latente permite asociar a cada encuesta un valor que en pruebas de educación correspondería a la habilidad del individuo. No obstante, para nosotros la consideramos como una valoración de la sección para la sala. Como tal valoración global, vamos a estudiar la correlación con la puntuación media que obtendría la encuesta como promedio de las respuestas. Estudiamos la correlación entre la VL y esa puntuación para todas las encuestas. Pero también estudiamos las medias por sala que se obtienen por las dos formas. A continuación, mostramos los resultados obtenidos en cada sección.

2.1. Sección A: Aspectos generales

En la siguiente gráfica vemos la distribución de las puntuaciones en la variable latente de todos los encuestados, para cada librería. Se muestra una tabla con las correlaciones entre la variable latente para cada encuesta y la puntuación global obtenida.

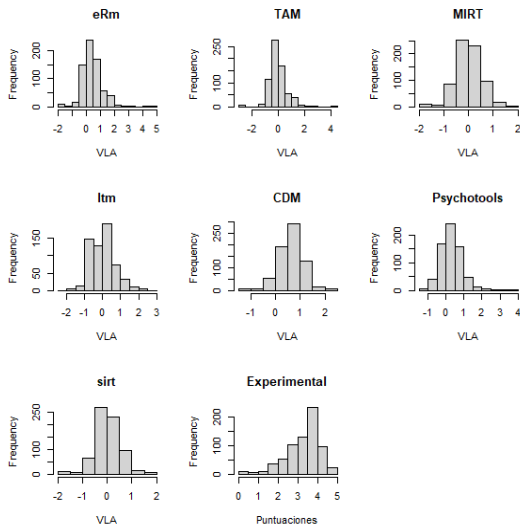


Figura 1. Distribución variable latente sección A para las diferentes librerías y distribución puntuaciones.

Tabla 1. Correlaciones entre la variable latente para la sección A entre todas las librerías.

	tam	mirt	ltm	cdm	psy	sirt	Experimental
erm	0.998	0.959	-0.022	0.956	1.000	0.955	0.876
tam		0.955	-0.020	0.953	1.000	0.952	0.875
mirt			-0.037	0.998	0.973	0.997	0.972
ltm				-0.032	-0.013	-0.042	-0.049
cdm					0.970	0.989	0.969
psy						0.968	0.901
sirt							0.969

En la siguiente tabla consideramos las correlaciones entre la VL y las puntuaciones por sala.

Tabla 2. Correlaciones entre la variable latente para la sección A y las puntuaciones por salas.

	tam	mirt	ltm	cdm	psy	sirt	Experimental
erm	1.000	0.985	-0.191	0.986	0.972	0.984	0.948
tam		0.982	-0.181	0.982	0.966	0.981	0.942
mirt			-0.193	1.000	0.985	0.999	0.987
ltm				-0.176	-0.220	-0.209	-0.185
cdm					0.984	0.998	0.985
psy						0.987	0.964
sirt							0.987

Finalmente estudiamos el error cuadrático medio entre la VL y la puntuación por sala. Como la Variable latente está definida en un intervalo de valores diferente al utilizado en las puntuaciones de cada pregunta, se ha optado por centrar cada distribución en su media. Es decir, se ha calculado

$$RMSE = \sqrt{\frac{(\hat{x} - \bar{\hat{x}}) - (x - \bar{x})^2}{N}}$$

Tabla 3. Error cuadrático medio entre VLA y puntuaciones por sala.

Librería	RMSE
erm	0.108
tam	0.115
mirt	0.142
ltm	0.551
cdm	0.141
psy	0.129
sirt	0.142

2.2. Sección B: Percepción acústica detallada

En la siguiente gráfica vemos la distribución de las puntuaciones en la variable latente para la sección B de todos los encuestados, para cada librería. Se muestra una tabla con las correlaciones entre la variable latente para cada encuesta y la puntuación global obtenida.

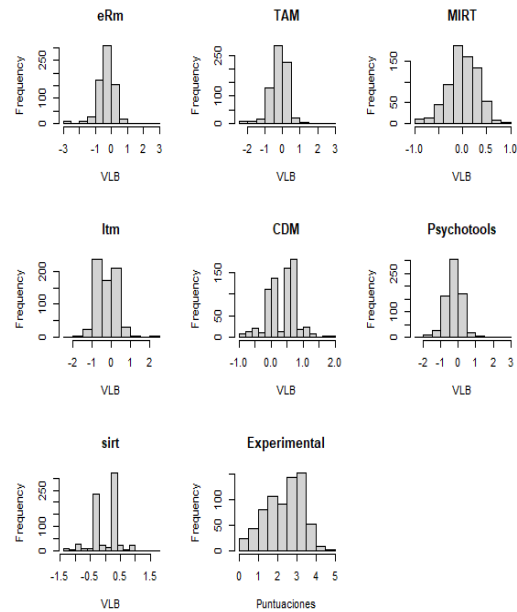


Figura 2. Distribución variable latente sección B para las diferentes librerías y distribución puntuaciones.

Y finalmente los RSME trasladados para la sección C.

Tabla 9. Error cuadrático medio entre VLC y puntuaciones por sala.

Librería	RMSE
erm	0.175
tam	0.184
mirt	0.165
ltm	0.449
cdm	0.222
psy	0.206
sirt	0.22

3. DISCUSIÓN

Dados los resultados obtenidos, nos decantamos por utilizar la función *PCM* de la librería *eRm*. En las siguientes gráficas podemos ver la relación entre la variable latente y las puntuaciones medias por sala para esta librería.

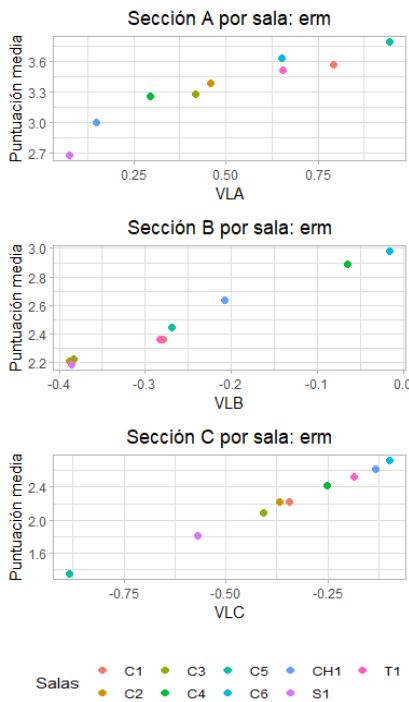


Figura 4. Dispersión entre las puntuaciones y la variable latente para la librería “*eRm*” por sección y salas.

4. CONCLUSIONES

Como hemos mostrado en los resultados para cada sección, la variable latente que proporciona un análisis TRI mediante el modelo de crédito parcial de las encuestas de valoración acústica de salas correlaciona con las puntuaciones medias. Las gráficas muestran que la distribución de la VL no es exactamente igual que las puntuaciones experimentales. Pero las correlaciones entre los valores medios en cada sala de las VL y sus puntuaciones son muy buenas. Si consideramos los errores cuadráticos medios, hemos visto que si bien para algunas librerías no hay muchas diferencias, el mejor resultado para la sección A y B se obtiene con la librería “*erm*”. Para la sección C con la librería “*mirt*”.

La determinación de la variable latente mediante TRI tiene una ventaja sobre la puntuación media directa o experimental, puesto que estas técnicas se han preocupado por obtener las habilidades reales de los individuos estudiados. De tal manera que el análisis de todas las encuestas permite establecer la VL de cada una de ellas con garantía. Estos resultados se han presentado en un reciente congreso [12].

Nuestro siguiente objetivo es estudiar las relaciones entre la variable latente que proporciona este método de análisis y los parámetros objetivos que se utilizan habitualmente para cuantificar la acústica de las salas. También creemos que será interesante estudiar si existe alguna relación entre la variable latente aquí obtenida y los factores que obtuvimos cuando realizamos estudios de componentes principales a estas mismas encuestas [6].

5. REFERENCIAS

[1] Andrich, D., and I. Marais. 2019. *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Springer Texts in Education. Springer Nature Singapore. <https://books.google.es/books?id=7dKiDwAAQB AJ>.

[2] Baker, & Kim, F. B. 2004. *Item Response Theory: Parameter Estimation Techniques*. 2nd ed. CRC Press.



- [3] Chalmers, R. Philip. 2012. "Mirt: A Multidimensional Item Response Theory Package for the r Environment." *Journal of Statistical Software* 48 (6): 1–29. <https://doi.org/10.18637/jss.v048.i06>.
- [4] George, A., A. C. AND Robitzsch. 2015. "Cognitive Diagnosis Models in r: A Didactic." *The Quantitative Methods for Psychology* 11 (3): 189–205. <https://doi.org/10.20982/tqmp.11.3.p189>.
- [5] Giménez, Alicia, Rosa M. Cibrián, Sara Girón, Teófilo Zamarreño, Juan José Sendra, Ana Vela, and F. Daumal. 2011. "Questionnaire Survey to Qualify the Acoustics of Spanish Concert Halls." *Acta Acustica United With Acustica* 97: 949–65. <https://api.semanticscholar.org/CorpusID:111423124>.
- [6] Giménez, Alicia, Rosa Cibrián, and Salvador Cerdá. 2013. "Subjective Assessment of Concert Halls: A Common Vocabulary for Music Lovers and Acousticians." *Archives of Acoustics* 37 (3). <https://acoustics.ippt.gov.pl/index.php/aa/article/view/98>.
- [7] Mair, Patrick, and Reinhold Hatzinger. 2007. "Extended Rasch Modeling: The eRm Package for the Application of IRT Models in r." *Journal of Statistical Software* 20 (9): 1–20. <https://doi.org/10.18637/jss.v020.i09>.
- [8] Rizopoulos, Dimitris. 2006. "Ltm: An r Package for Latent Variable Modeling and Item Response Analysis." *Journal of Statistical Software* 17 (5): 1–25. <https://doi.org/10.18637/jss.v017.i05>.
- [9] Robitzsch, Alexander. 2023. *Sirt: Supplementary Item Response Theory Models*. <https://CRAN.R-project.org/package=sirt>.
- [10] Robitzsch, Alexander, Thomas Kiefer, and Margaret Wu. 2022. *TAM: Test Analysis Modules*. <https://CRAN.R-project.org/package=TAM>.
- [11] Schneider, Lennart, Carolin Strobl, Achim Zeileis, and Rudolf Debelak. 2022. "An R Toolbox for Score-Based Measurement Invariance Tests in IRT Models." *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01689-0>.
- [12] Salvador Cerdá, Carlos Abellán, Jaime Segura, Alicia Giménez, Miguel Arana, Rosa M. Cibrián. Use of item response theory (IRT) in subjective assessment of concert halls. *Internoise* 2023.