

## Introducción Sección Especial

### El Modelo de Rasch: Nuevos desarrollos y aplicaciones en la ciencia de la medición objetiva

<sup>1</sup>Brent Duckor, <sup>2</sup>María Verónica Santelices y <sup>3</sup>Steffen Brandt

<sup>1</sup> College of Education, San Jose State University, California, EE.UU.

<sup>2</sup>Facultad de Educación, Pontificia Universidad Católica de Chile

<sup>3</sup>Art of Education, Statistical Analyses

Hace más de 50 años, Georg Rasch contribuyó a la fundación del campo de la Teoría de Respuesta al Ítem con el modelo que lleva su nombre, caracterizada por el uso de un solo parámetro para modelar la relación entre la dificultad del ítem y la habilidad de una persona (Rasch, 1960/1980, 1977). Considerado durante mucho tiempo como el estándar de excelencia en la medición «objetiva», se han propuesto, aplicado e investigado numerosas extensiones del modelo de Rasch, para su utilización en pruebas educacionales y evaluación educacional. Según el Comité de Programa del Instituto de Medición Objetiva (2000), la medición «objetiva» es «la repetición de una cantidad unitaria que mantiene su tamaño... independientemente de cuál sea el instrumento... utilizado e independiente de a quién o qué persona o cosa se mida». El Modelo de Escalas de Clasificación, el Modelo de Crédito Parcial y el Modelo Logístico Multinomial de Coeficiente Aleatorio, son parte de la familia Rasch. Enfocados en un principio en logros de lectura, los modelos de Rasch se emplean hoy en día en campos tan diversos como atención médica, fisioterapia, enfermería, farmacéutica y deportes competitivos.

El modelo de Rasch expone nuestra teoría de la estructura de cualquier variable latente a una rigurosa prueba científica al preguntar: «¿Los datos se ajustan al modelo?» (Wright & Master, 1980). Desde una perspectiva de elaboración de mediciones (Wilson, 2014), el marco del modelo de Rasch permite a los investigadores explorar la validez de las escalas, a partir de la construcción de una teoría y el contraste de dicha teoría con la evidencia del ordenamiento empírico de personas e ítems en un mapa de Wright. Muy en línea con la noción actual de validez como un argumento (Kane, 2015), el marco de Rasch ayuda a explorar la estructura interna del constructo y la validez de su contenido, ambos al mismo tiempo.

Además, Wilson (2005) nos recuerda que el marco de Rasch propone una serie de principios básicos para la medición de personas y la calibración de ítems en una sola escala; lejos de ser solo otra técnica de cuantificación, este enfoque es fundamentalmente importante para las interpretaciones realizadas en la ciencia de la medición (Duckor, Draney, & Wilson, 2009). Para escoger y evaluar un modelo de medición, los investigadores deben pensar en forma espacial, en términos de un mapa geográfico. La idea de la «ubicación» de una respuesta al ítem con respecto a la ubicación de otra respuesta al ítem solo tiene sentido si dicho significado relativo es independiente a la ubicación del encuestado implicado. Es decir, la interpretación de ubicaciones relativas debe ser uniforme e independiente al lugar donde se encuentre el encuestado. Este requerimiento de invariabilidad corresponde a la idea de que una «pulgada representa una milla» o un «metro representa un kilómetro», en cualquier lugar de un mapa geográfico.

Durante años, el modelo uniparamétrico (Rasch) ha enfrentado objeciones de los investigadores de la Teoría de la Respuesta al Ítem (TIR), que favorecen modelos más flexibles y de mejor ajuste a los datos. Algunos de los investigadores de la TIR (Birnbaum, 1968) han impulsado la inclusión de parámetros adicionales con el fin de, por ejemplo, ajustar la variación en discriminaciones de ítem (2PL) o la variación en probabilidades de adivinación (3PL). Mientras el modelo de Rasch requiere que todos los ítems discriminen de igual manera para poder definir la habilidad que se está midiendo, los modelos 2PL y 3PL permiten que la discriminación varíe entre ítems y se calcula recursivamente como parte del proceso de estimación. Esto también impacta la estimación del parámetro de dificultad del ítem, creando una acentuada diferencia en la forma en que las dificultades de los ítems de Rasch y 2PL/3PL se interpretan. Debido a que la discriminación de ítem es, al menos en parte, una propiedad de cómo una muestra particular de examinados interactúa con un ítem y no es exclusivamente una propiedad del ítem, y porque los examinados varían entre las pruebas, la consecuencia ineludible es que las puntuaciones calculadas al usar los modelos 2PL y 3PL no puede garantizarse que las puntuaciones sean generalizables entre pruebas tales como aquellas calculadas a partir de datos que están restringidos para ajustarse a las exigencias del modelo de Rasch.

Así, aunque los modelos 2PL y 3PL podrían ajustarse mejor a los datos, el beneficio frecuentemente tiene un precio. Primero, la interpretación de las puntuaciones de las pruebas no es tan clara, puesto que los diferentes parámetros de discriminación de los ítems se traducen en distintas ponderaciones de los ítems en la puntuación final de la prueba, sin poder distinguir si esto se debe a una característica de un ítem o a una propiedad de la muestra. Segundo, el principio científico de poder verificar nuestras expectativas teóricas en relación a un constructo (en un marco de validez), se pierde si los investigadores no pueden falsificar sus hipótesis sobre la estructura que tendrán los datos. Además, estudios recientes (San Martín, González, & Tuerlinckx, 2015) han demostrado la no identificabilidad del modelo 3PL, examinando cómo, incluso después de fijar la dificultad, la discriminación y los parámetros de adivinación de un ítem, los parámetros de los ítems restantes no pueden identificarse utilizando las observaciones disponibles, no poseen una interpretación empírica y no pueden estimarse de forma imparcial y consistente.

Estas razones explican por qué el modelo de Rasch aún se utiliza hoy en día y se mantiene como una herramienta importante entre los instrumentos de medición de los psicometristas y expertos en evaluación educacional.

De acuerdo con las tradiciones establecidas por las series precedentes, *Medición Objetiva: Teoría en Práctica (Vols. 1-5)* y *Avances en Medición de Rasch (Vols. 1-2)*, esta publicación se complace en presentar desarrollos tanto teóricos como prácticos de los modelos de medición de Rasch. Todos los papers incluidos en esta sección especial se presentaron originalmente en el Taller Internacional de Medición Objetiva 2014 (International Objective Measurement Workshop [IOMW]) en Filadelfia, Pensilvania, Estados Unidos. El IOMW es una conferencia bianual que precede a la conferencia de la Asociación Americana de Investigación Educativa (American Educational Research Association, AERA) y reúne a expertos de todo el mundo para compartir sus trabajos en las áreas de modelamiento usando el modelo

de Rasch, psicometría y filosofía de la medición. Los manuscritos se solicitaron para mantener cierta coherencia temática, y cada texto fue sometido a una revisión ciega por al menos dos expertos. Los cinco artículos incluidos en esta sección especial «El Modelo de Rasch: Nuevos Desarrollos y Aplicaciones en la Ciencia de la Medición Objetiva», consideran el modelo de Rasch desde distintos ángulos y aplicaciones.

El primer artículo de Andrich (2015), «Componentes de la varianza de escalas con una estructura de subescalas empleando dos cálculos de coeficiente  $\alpha$ », aborda la problemática del análisis de dimensionalidad, el coeficiente alfa y la varianza de subescalas. Señala que las escalas construidas para medir una variable individual están compuestas por subescalas de ítems, que miden diferentes aspectos de la variable. Utilizando un estudio de simulación, el paper propone un método simple que se puede utilizar para proporcionar un resumen más exhaustivo de las propiedades de una escala con subescalas de aquel que es posible usando solo una estimación del coeficiente de confiabilidad. Demuestra que, con algunas hipótesis de simplificación y utilizando una estructura bifactorial, la razón de dos cálculos del coeficiente alfa, uno al nivel de los ítems y el otro al nivel de las subescalas, se puede usar para obtener (a) la proporción de varianza común verdadera, (b) la proporción de varianza única verdadera, (c) la proporción de varianza común verdadera relativa a la suma de las varianzas común verdadera y única, y (d) una correlación «resumen» entre subescalas corregida inmediatamente para la atenuación causada por el error. El artículo muestra cómo los especialistas en el modelo de Rasch pueden entregar nuevas luces en problemas tradicionales y planteamientos confiables para quienes utilizan el coeficiente alfa.

El segundo artículo, realizado por Behizadeh y Engelhard (2015), «Evaluación válida de la escritura desde la perspectiva de las comunidades de investigación en escritura y medición», aborda problemas relacionados con la validez, validez consecuencial, evaluación de escritura y comunidades de práctica, con un foco particular en la teoría de medición de Rasch. El estudio analiza el concepto de validez en dos comunidades de práctica distintas: las comunidades de la investigación de la escritura y las comunidades de mediciones educativas. Destaca las contribuciones que la *teoría de medición de Rasch* (Rasch, 1960/1980) aporta para comprender y evaluar la validez. Al relacionar perspectivas técnicas y no técnicas sobre validez, los autores exploran puntos de consenso y convergencia en relación a la validez. Dicho estudio tiene implicancias para mejorar la investigación, la teoría y la práctica de la evaluación de la escritura tanto para académicos como para profesionales.

El tercer artículo, elaborado por Fisher y Wilson (2015), «Construyendo una zona de intercambio productiva en la investigación y práctica de la evaluación educacional», explora los desafíos de la medición a través de diferentes contextos institucionales. Sostiene que diversos puntos de vista relacionados al acto de medir —desde el laboratorio de ciencias a la sala de clases al mercado— se pueden reconciliar con el uso de objetos de demarcación que admiten significados comunes. Los autores describen cómo ejemplos modelados psicométricamente, conocidos como *mapas de constructos* y *mapas Wright*, desarrollados en base al modelo de Rasch, funcionan como objetos de demarcación y pueden servir de base para analogías productivas en evaluación educativa a través de (a) la preservación de estructuras relacionales, (b) la realización de mapeos isomórficos entre sistemas y (c) el fomento de la sistematización, entendida como sistemas de mapeo de estructuras relacionales de orden superior. A través del uso del caso de BEAR Assessment System y su software adjunto, el artículo explora cómo dichas tecnologías sustentan las alianzas prácticas entre enseñanza, formulación de políticas, desarrollo de evaluación y currículum, psicometría y tecnología de la información.

En el cuarto artículo, Korpershoek (2015) presenta «Estudio de confiabilidad y validez de la Escala de Manejo de Compromisos Identitarios de Utrecht, adaptada para medir procesos de formación de identidad de estudiantes en su universidad», que aborda los problemas de desarrollo del compromiso y la identidad en mediciones psicológicas. El artículo examina evidencia para la validez del constructo y la validez predictiva del marco de medición presentado por Crocetti, Rubini y Meeus (2008), desarrollado originalmente para cuantificar los procesos de la formación de la identidad (como el logro de compromisos contraídos) de individuos en diversos ámbitos. Usando un grupo de estudiantes universitarios y utilizando un modelo multidimensional de Rasch, el artículo analiza una versión adaptada de la Escala de Manejo de Identidad de Utrecht (U-MICS) para medir los procesos de formación de identidad de estudiantes universitarios, considerados parte de las identidades personales de los estudiantes en su universidad. Los resultados demostraron algunos signos de que la estructura interna puede representar los procesos de formación de la identidad de los estudiantes en la universidad, aunque ciertos ítems requieren mejoras adicionales para ajustarse al modelo multidimensional.

El quinto artículo, realizado por Williamson (2015), «La medición del crecimiento académico contextualiza la complejidad del texto», aborda los desafíos y oportunidades para medir el desarrollo de la lectura. Al utilizar los principios fundamentales de Rasch para la medición objetiva específica, el artículo argumenta que la medida óptima del desarrollo académico de los estudiantes requiere una escala que sea unidimensional, continua, de intervalos iguales, evolutiva e invariable respecto a la ubicación y el tamaño de la unidad. Proporciona un ejemplo de la medición de Rasch validado empíricamente y que posee una teoría de constructo de lectura operacionalizado. Al utilizar la escala Lexile, el artículo ilustra un continuum de la complejidad del texto, en donde personas (lectores) e ítems (textos) se incluyen en una escala de «desarrollo académico» común. Además, el artículo explora avances recientes en políticas educativas (Estándares de Preparación para las Profesiones y la Universidad) en Estados Unidos que recomiendan una mayor exposición de los estudiantes a textos complejos. El autor advierte que el modelado paramétrico de curvas de crecimiento alternativo puede enmarcar de forma concreta y tangible las conversaciones sobre cómo los estudiantes podrían lograr dichas metas de preparación para las profesiones y para la universidad.

Estos cinco artículos son ejemplos notables de la combinación del trabajo teórico y empírico que caracteriza a los análisis Rasch, y esperamos que motiven a los lectores a aprender más sobre los modelos de Rasch y sus usos potenciales en la elaboración de argumentos de validez basados en evidencias. Los artículos muestran cómo los conceptos del modelo de Rasch y su modelado estadístico son aplicables a una amplia gama de temas, que incluyen el desarrollo de instrumentos, programas de investigación innovadora y políticas públicas. Esperamos también que esta sección especial motive a los lectores a participar en la próxima conferencia IOMW que se realizará en Washington, D.C. en abril de 2016.

---

**Referencias**

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test punctuations* (pp. 395-479). Reading, MA: Addison-Wesley.
- Crocetti, E., Rubini, M., & Meeus, W. (2008). Capturing the dynamics of identity formation in various ethnic groups. Development and validation of a three-dimensional model. *Journal of Adolescence*, *31*, 207-222. doi: 10.1016/j.adolescence.2007.09.002
- Duckor, B., Draney, K., & Wilson, M. (2009). Measuring measuring: Toward a theory of proficiency with the Constructing Measures framework. *Journal of Applied Measurement*, *10*(3), 296-319.
- Kane, M. T. (2015). Explicating validity. *Assessment in Education: Principles, Policy & Practice*. Recuperado de <http://dx.doi.org/10.1080/0969594X.2015.1060192>
- Program Committee of the Objective Measurement Institute (2000). *Definition of objective measurement*. Recuperado de <http://www.rasch.org/define.htm>
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press.
- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy*, *14*, 58-93.
- San Martín, E., González, J., & Tuerlinckx, F. (2015). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika*, *80*(2), 450-467. doi: 10.1007/s11336-014-9404-2
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wilson, M. (2014). Considerations for measuring learning progressions where target learning is represented as a cycle. *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, *51*(1), 156-174. doi: 10.7764/PEL.51.1.2014.12
- Wright, B. D., & Masters, G. N. (1981). *The measurement of knowledge and attitude*. Chicago: Statistical Laboratory, Department of Education, University of Chicago.