

CUARTO INFORME DEL ESTADO DE LA EDUCACIÓN

Aporte Especial

El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala

Investigadoras:

Shirley Rojas

Eiliana Montero

2012



CONSEJO NACIONAL DE RECTORES

Nota: Las cifras de las ponencias pueden no coincidir con las consignadas por el IV Informe Estado de la Educación en el tema respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala

Las pruebas estandarizadas son un recurso utilizado en el ámbito educativo para medir constructos complejos. Tal y como se explica en el cuerpo del capítulo especial, las pruebas estandarizadas corresponden a instrumentos de medición en cuya construcción y validación se utiliza un enfoque científico que permita obtener evidencias de la validez y confiabilidad de las mismas. El objetivo de esta sección es introducir a los lectores interesados al modelo de medición más utilizado actualmente en pruebas estandarizadas educativas de gran escala.

Durante muchos años los educadores y los evaluadores educativos han cuestionado a la psicometría en cuanto a su utilidad en pruebas educativas, dado que los puntajes generados a partir del modelo de medición más antiguo y conocido de la Teoría Clásica de los Test (TCT), no proveen información acerca de qué puede o no hacer el estudiante. Por ejemplo, el porcentaje de respuestas correctas o el puntaje promedio en un examen, no dice nada per se acerca del desempeño del examinado, no brinda información sobre cuáles contenidos domina y cuáles no, y en qué nivel se encuentra con respecto a los contenidos mínimos que debería dominar. Si un estudiante obtuvo una nota de 80 en Matemáticas no se puede establecer el grado de dominio de los contenidos o procesos que están presentes en el examen. Además, en la TCT se supone que el error de medición asociado a cada puntaje observado es igual para todos los examinados. Adicionalmente, a partir de esta teoría se produce un resultado conocido como el coeficiente de confiabilidad alfa de Cronbach, usado para medir la confiabilidad de las pruebas.

La TCT fue diseñada en el contexto de los estudios iniciales sobre la inteligencia y en estas circunstancias generó un modelo interpretativo que se denomina de referencia a normas, y es hasta el día de hoy el que se utiliza para interpretar las llamadas pruebas tradicionales de inteligencia. Así, el significado que se otorga a las puntuaciones de una prueba bajo este modelo se define en torno a la posición relativa que ocupa un examinado particular en relación con el grupo total de examinados que tomó el test; es usual hacer interpretación en términos de percentiles y también en torno a cuántas unidades por encima o por debajo del promedio se ubica el sujeto. Sin embargo, en un contexto educativo es usualmente necesario, además de establecer la posición relativa del examinado, ser capaz de utilizar las puntuaciones de las pruebas para determinar a nivel absoluto las fortalezas y debilidades del individuo en torno al constructo que interesa medir; en este sentido hablamos de interpretaciones referidas a criterios. Por tanto, como se explicó anteriormente, la TCT no cumple a cabalidad con este propósito.

En respuesta a la evidencia de las limitaciones de la TCT (Embretson & Hershberger, 1999) han surgido nuevos modelos de medición. Tal es el caso del modelo de Rasch, un tipo particular de los modelos de Teoría de Respuesta a los Ítems (TRI), el cual tiene propiedades únicas y muy llamativas que permiten obtener interpretaciones con respecto a criterios en pruebas educativas. El modelo de Rasch supera al modelo TCT por distintas razones. Como una de las razones se puede mencionar la gran distinción entre ambos en cuanto a la estimación del atributo o constructo que se pretende medir

con las pruebas. Con la TCT se obtiene un puntaje con la suma de las respuestas a los ítems, mientras que en el caso de Rasch, por medio del patrón de respuesta de cada estudiante se logra estimar el nivel en el constructo medido. Además, en el modelo de Rasch se establece una medición conjunta debido a que se posiciona en la misma escala la dificultad de los ítems y el nivel en el constructo de los examinados, lo cual no puede establecerse con la TCT.

El modelo de Rasch presenta ventajas fundamentales a nivel interpretativo y práctico para el análisis de los resultados de pruebas educativas. Gracias a estas ventajas el modelo de Rasch se ha vuelto muy popular entre los programas de pruebas internacionales y en general también a nivel de los programas nacionales de evaluación educativa que utilizan pruebas estandarizadas. En particular los beneficios del modelo de Rasch para el análisis de pruebas educativas pueden aplicarse a las pruebas PISA, a las pruebas de diagnóstico o bien a pruebas de certificación (como en el caso de las pruebas de Bachillerato). Estas ventajas fundamentales corresponden a las siguientes:

- A partir de la puntuación obtenida en Rasch se puede establecer a nivel sustantivo los logros del estudiante en cuanto a procesos y contenidos presentes en el examen y también sus áreas de oportunidad en términos del constructo objeto de medición.
- Se genera una estimación del error para cada puntuación estimada en el constructo de interés. Esto permite generar pruebas de mayor calidad puesto que se tienen herramientas objetivas para poder adaptar las pruebas a los propósitos de interés y a las poblaciones metas a las que van dirigidas. Por ejemplo, una prueba de certificación académica como los exámenes nacionales de Bachillerato puede utilizar el modelo de Rasch para minimizar el error de medición en el punto de corte de la escala que define al grupo de estudiantes que merecen obtener la certificación, lográndose así decisiones más válidas y confiables en términos de quiénes merecen promover y quiénes no.
- Rasch permite no solamente juzgar el ajuste de los ítems en términos de sus propiedades psicométricas sino que también permite establecer a nivel de los examinados si las respuestas brindadas en el examen sugieren posibilidad de respuestas al azar o con evidente descuido, así como la posibilidad de copia, información que a nivel investigativo y operativo puede ser de gran interés en torno a las posibles causas de estas situaciones y las formas de evitarlas.
- En el modelo de Rasch se cumple que diferencias iguales en la puntuación representan diferencias iguales en el constructo. Esto se puede ejemplificar de la siguiente manera: se tienen dos alumnos Juan y María, cuya nota en una prueba fue respectivamente 20 y 40 (en una escala de 0 a 100); mientras que otros dos alumnos, Sandra y Pablo, obtuvieron en la misma prueba 80 y 100 respectivamente. Entre cada par de alumnos la diferencia es de 20 puntos, por lo que la diferencia en el constructo en cada uno de los dos casos también es de 20. Es decir, diferencias del mismo tamaño en las puntuaciones (en este ejemplo, de tamaño 20), representan diferencias del mismo tamaño en el constructo (también 20).

- El modelo de Rasch siempre es estimable, a diferencia de los otros modelos TRI. Es una gran ventaja a nivel de los programas internacionales que trabajan con un enfoque de toma de decisiones y no investigativo, porque se aseguran de que siempre van a poder estimar el modelo y obtener resultados.

A continuación se hace una breve descripción del modelo, y de las propiedades únicas que posee que permiten entonces obtener estas ventajas a nivel práctico y de interpretación. Además se utilizan resultados de PISA 2009 de Costa Rica para ejemplificar el tipo de análisis y ventajas que permite este modelo.

El modelo de Rasch

El modelo de Rasch es un recurso para analizar la calidad técnica de ítems que forman parte de una prueba, tal y como las pruebas educativas de competencias de PISA. Es por esto que el programa “PISA” utiliza ampliamente el análisis con el modelo de Rasch como parte de sus análisis estadísticos de las pruebas (OECD, 2012).

Su fórmula matemática relaciona la probabilidad de responder el ítem correctamente, la diferencia entre el nivel en el constructo o atributo del individuo y la dificultad del ítem (Rasch, 1960). Si el nivel del examinado en el atributo que se desea medir es mayor que la dificultad del ítem entonces la probabilidad de responder correctamente el ítem es mayor que la probabilidad de responderlo incorrectamente. Se establece la relación inversa en el caso de que el nivel del examinado en el atributo que se desea medir sea menor que la dificultad del ítem, es decir, la probabilidad de responder correctamente al ítem va a ser menor que la probabilidad de no responderlo correctamente. Lo anterior se ilustra más adelante en “Dificultad de los ítems y nivel en el atributo de los examinados”.

Propiedades del modelo de Rasch

Las propiedades únicas que posee el modelo de Rasch a nivel matemático y estadístico son las que permiten obtener las ventajas a nivel aplicado e interpretativo que se listaron con anterioridad. Las propiedades clave del modelo Rasch según Prieto y Delgado (2003) son las siguientes:

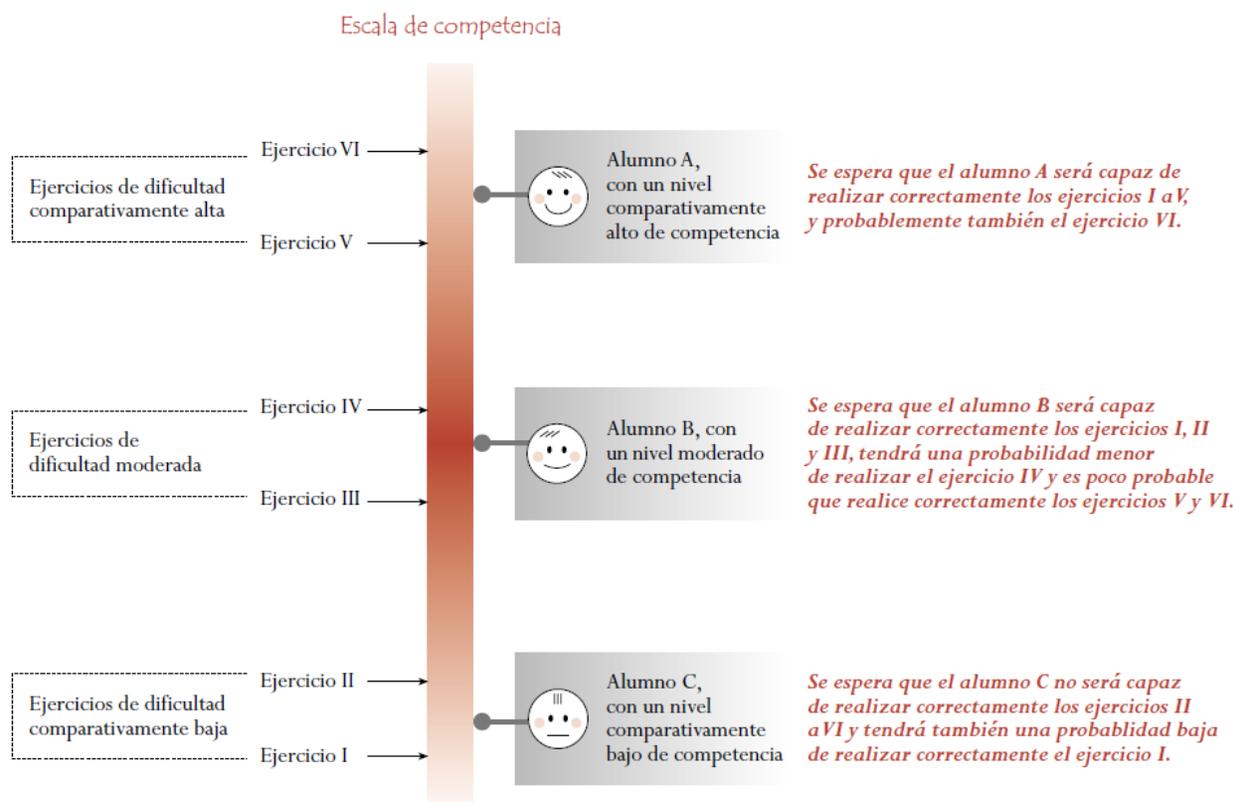
- *Medición conjunta*: permite ubicar en la misma escala tanto la dificultad de los ítems como el nivel de del atributo que se desea medir de los examinados ya que ambos parámetros tienen las mismas unidades, permitiendo establecer la relación entre el nivel del constructo que representan los ítems y el nivel del constructo que los estudiantes poseen. De aquí se deriva el valor diagnóstico del modelo de Rasch, valor particularmente útil en pruebas como las de PISA, debido a que esta propiedad permite la obtención de los niveles de desempeño, que son los criterios con los que se comparan los resultados de los examinados.

- *Objetividad específica:* las medidas del nivel del atributo que se desea medir en los examinados y de la de la dificultad de los ítems no dependen de las condiciones específicas en que han sido obtenidas. Por lo tanto la diferencia de dos examinados en el atributo no depende de los ítems y la diferencia entre las dificultades de los ítems no depende de las personas que se evaluaron para obtener esas dificultades. Esto representa una propiedad importante del modelo de Rasch ya que en la TCT y en otros modelos TRI más conocidos, la dificultad de los ítems puede variar dependiendo del grupo de examinados y la estimación del nivel del atributo de los examinados puede cambiar dependiendo del grupo de ítems que le fueron administrados.
- *Propiedades de intervalo:* la interpretación de las diferencias es la misma a lo largo de la escala del atributo o constructo que se mide. Las diferencias constantes entre ítems y examinados tienen asociada la misma probabilidad de responder correctamente. Con esto se supera la limitación de la TCT, donde se tienen escalas ordinales y por ende no se puede determinar que las diferencias son las mismas.
- *Especificidad del error típico de medida:* se evidencia que el test no tiene la misma fiabilidad a lo largo de todo el atributo, y permite diferentes errores en los distintos niveles, es decir, distinta precisión en la medición en cada uno de los niveles del atributo. En la TCT se asume que se mide con la misma fiabilidad a lo largo del atributo, aunque se sabe que esto no sucede así.

Dificultad de los ítems y nivel en el atributo de los examinados

La característica de medición conjunta establece de forma muy clara la relación entre la dificultad de ítems y el nivel de los examinados. Dicha propiedad se puede representar de forma gráfica mediante una escala de competencia. Un ejemplo de dicha escala corresponde a un mapa presentado por la OECD (2004) en el Informe PISA 2003 “Aprender para el mundo del mañana”. En esta escala de competencia se sitúa los ítems, del más difícil al más fácil, y los examinados, del que tiene un nivel más alto de competencia al que tiene un nivel más bajo (Figura 1).

Figura 1
Relación entre los ítems del test y los estudiantes en una escala de competencia



En la figura se explica con gran claridad que entre más arriba de un ítem se encuentre el nivel del atributo o competencia de un examinado, es más probable que el examinado responda correctamente a dicho ítem. Y que por el contrario, entre más abajo de un ítem se sitúe el nivel del atributo de un examinado, es menos probable que dicho examinado pueda responder correctamente ese ítem.

Ilustración de la aplicación del modelo Rasch al análisis de resultados de la Pruebas PISA 2009 de Costa Rica

Un análisis aplicado del modelo de Rasch tomando como ejemplo ítems de dos cuadernillos de los que se aplicaron en Costa Rica en el 2009 se muestra en las siguientes figuras. Los ítems utilizados corresponden a los del Cuadernillo #8 de Matemática (Figura 2), y a los Cuadernillo #25 de Lectura (Figura 3). Este análisis es parte del proceso de PISA para el análisis de resultados, por lo que corresponde únicamente a una ilustración con una parte del conjunto de datos.

La información que proporcionan estos “mapas” permite clasificar los ítems de acuerdo a su dificultad y a los procesos y contenidos de los mismos; generando así divisiones en a escala que corresponden a los niveles de desempeño o competencia. Los niveles de desempeño para los dominios de Competencia Lectora y Matemática expuestos en el capítulo especial por Ana María Rodino y Óscar Salas, son ejemplos de los niveles de desempeño que se generan para las pruebas PISA con el modelo de Rasch.

Figura 2
Relación entre la competencia de los examinados y la dificultad de los ítems, Cuadernillo 8, ítems de Matemática
(Mapa de distribución de ítems vs personas)

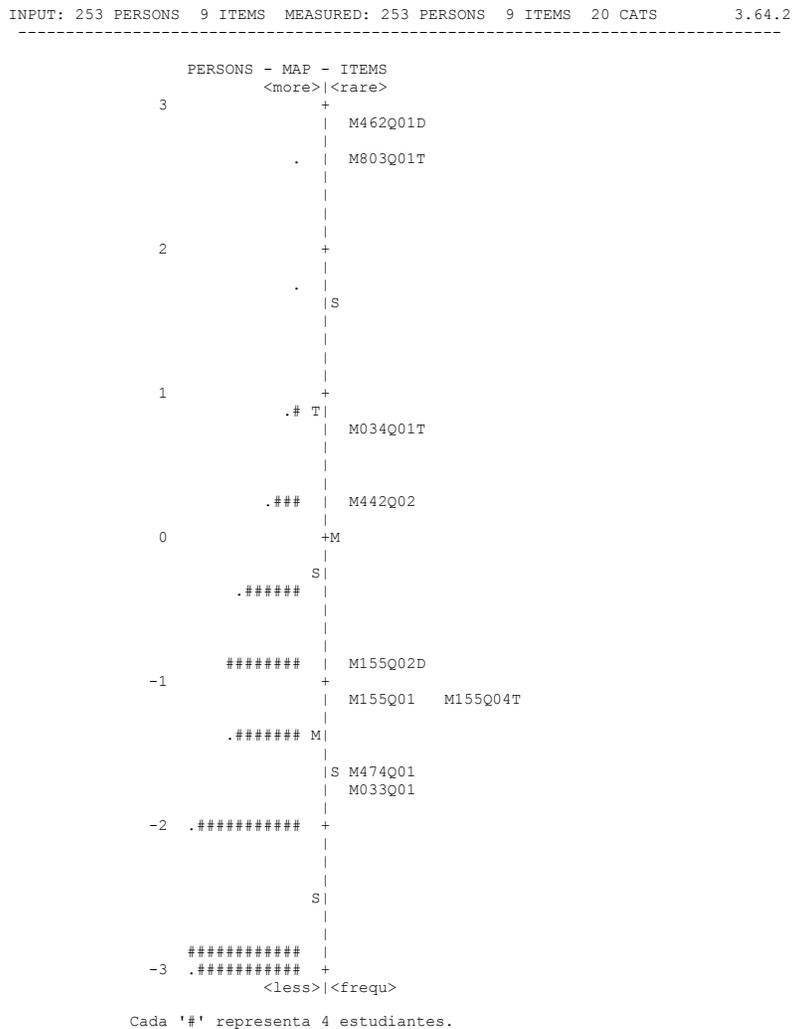
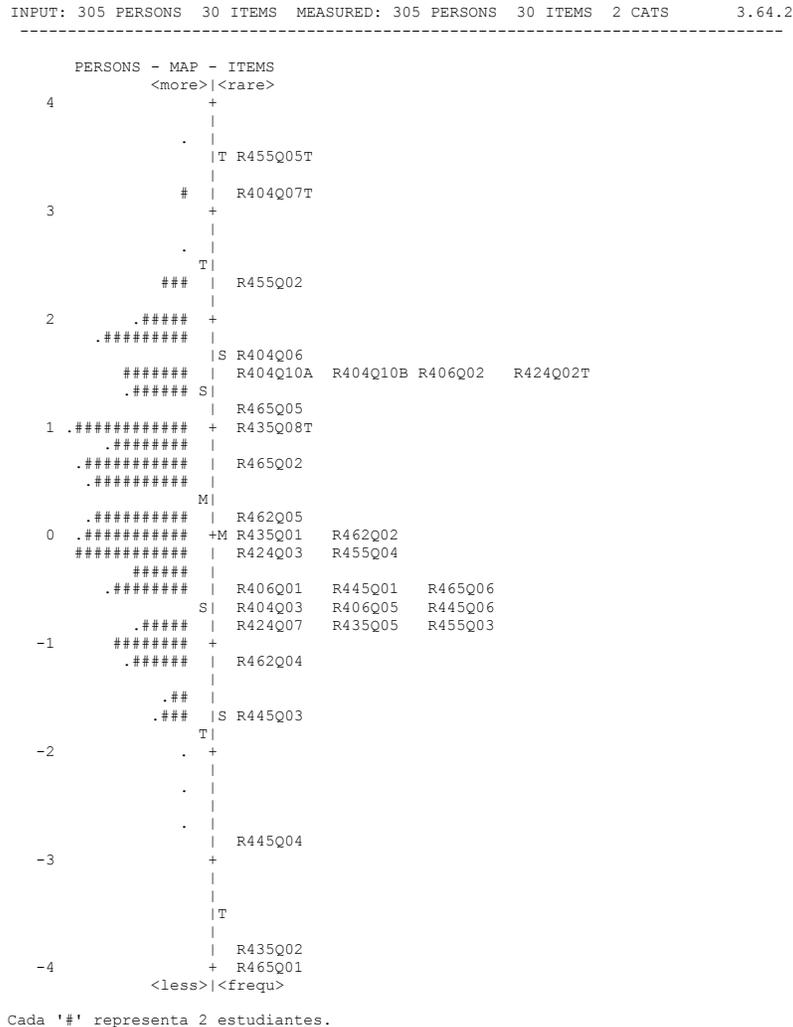


Figura 3
Relación entre la competencia de los examinados y la dificultad de los ítems, Cuadernillo 25, ítems de Lectura
(Mapa de distribución de ítems vs personas)



De la información que ofrecen los “mapas” presentados destaca lo siguiente:

- Los ítems del Cuadernillo #8 son en general más difíciles para ese grupo de estudiantes porque la media de los ítems (representada con la letra “M” a la derecha), está por encima de la media de los examinados (“M” a la izquierda).
- Asimismo se observa que en el Cuadernillo #8 hay una gran parte de los examinados (los que se ubican al final de la escala al lado izquierdo) para los cuales

no se puede generar información porque no tienen ítems asociados (del lado derecho); la dificultad y además la cantidad de ítems influyen en esta situación.

- Los ítems del cuadernillo #25 se distribuyen de tal forma que todos los examinados quedan representados por ítems, ya que donde se concentran los examinados también hay una concentración de los ítems.

El modelo de Rasch ha sido implementado en los sistemas de evaluación de educación alrededor del mundo, en programas con gran trayectoria y calidad. Los análisis que se realizan aplicando ese modelo permiten, como ya se ha mencionado, la presentación conjunta de la dificultad de los ítems y el nivel en el constructo de los examinados, y por ende los niveles de desempeño que se pueden generar a partir de ello. Asimismo, es particularmente útil en las situaciones en las que se requiera dar un diagnóstico individual. En general se justifica su aplicación en los sistemas de evaluación educativa, donde se puede utilizar este valor diagnóstico individual para formular recomendaciones acertadas con respecto a qué procesos y contenidos debe cambiar o mejorar, y tener una base para tomar esas decisiones más que con un simple puntaje. Existe la necesidad de que los evaluadores educativos en Costa Rica, y en nuestra región, conozcan y se familiaricen con estas herramientas que permiten generar interpretaciones muy útiles a nivel educativo.

Bibliografía

- Embretson, S.E. & Hershberger, S.L. (1999). *The new rules of measurement*. Mahwah, New Jersey: LEA.
- Jornet, J.M. & Suárez, J.M. (1996). Pruebas estandarizadas y evaluación del rendimiento: usos y características métricas. *Revista de Investigación Educativa*, 14(2): 141-163.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1): 57-66.
- OECD. (2004). *Informe PISA 2003: Aprender para el mundo del mañana*. España: Santillana Educación S.L.
- OECD. (2012). *PISA 2009 Technical Report*. PISA, OECD Publishing.
- Prieto, G. & Delgado A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.