

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.  
ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2026,  
Volumen 10, Número 1.

[https://doi.org/10.37811/cl\\_rcm.v10i1](https://doi.org/10.37811/cl_rcm.v10i1)

## **DESARROLLO DE INSTRUMENTOS DE MEDICIÓN UTILIZANDO MODELOS DE RASCH: ENSAYO METODOLÓGICO**

**DEVELOPMENT OF MEASUREMENT INSTRUMENTS USING  
RASCH MODELS: METHODOLOGICAL ESSAY**

**Jeremías Willmore Metivier**

Universidad Católica del Cibao-UCATECI, República Dominicana

DOI: [https://doi.org/10.37811/cl\\_rcm.v10i1.22699](https://doi.org/10.37811/cl_rcm.v10i1.22699)

## Desarrollo de Instrumentos de Medición utilizando Modelos de Rasch: Ensayo Metodológico

Jeremías Willmore Metivier <sup>1</sup>

[jeremiaswillmore@gmail.com](mailto:jeremiaswillmore@gmail.com)

<https://orcid.org/0000-0002-2759-2830>

Universidad Católica del Cibao-UCATECI  
República Dominicana

### RESUMEN

El desarrollo de instrumentos de medición requiere modelos que garanticen precisión, confiabilidad e invariancia en las mediciones. El modelo de Rasch, enmarcado dentro de la Teoría de Respuesta al Ítem (IRT), permite construir escalas unidimensionales y de intervalo, superando limitaciones propias de la Teoría Clásica de la Medición. El presente trabajo se desarrolla como un **ensayo metodológico**, cuyo propósito es analizar los fundamentos del modelo de Rasch, sus ventajas frente a otros enfoques y su utilidad en el proceso de calibración de ítems y medición de habilidades. **A manera de ejemplo ilustrativo, se emplean datos simulados, no provenientes de una aplicación empírica**, con el fin de mostrar el procedimiento general de análisis Rasch. Se describen modelos para ítems dicotómicos y politómicos, discutiendo criterios de ajuste, confiabilidad e interpretación mediante mapas ítem–sujeto. Finalmente, se resalta la importancia del modelo de Rasch como marco metodológico para la construcción de evaluaciones educativas más precisas y equitativas.

**Palabras clave:** confiabilidad, invariancia, calibración de ítems, modelo de Rasch, teoría de Respuesta al Ítem

---

<sup>1</sup> Autor principal

Correspondencia: [jeremiaswillmore@gmail.com](mailto:jeremiaswillmore@gmail.com)

# Development of Measurement Instruments using Rasch Models: Methodological Essay

## ABSTRACT

The development of measurement instruments requires models that ensure precision, reliability, and invariance in measurement. The Rasch model, framed within Item Response Theory (IRT), enables the construction of unidimensional, interval-level scales, thereby overcoming key limitations of Classical Test Theory. **This paper is presented as a methodological essay**, aimed at analyzing the theoretical foundations of the Rasch model, its advantages over alternative approaches, and its utility in item calibration and ability measurement. **For illustrative purposes, simulated data—rather than data from an empirical application—are employed** to demonstrate the general Rasch analysis procedure. Models for dichotomous and polytomous items are described, with discussion of fit criteria, reliability, and interpretation through item–person maps. Finally, the relevance of the Rasch model as a methodological framework for the development of more precise and equitable educational assessments is highlighted.

**Keywords:** Rasch model, Item Response Theory, item calibration, educational measurement, Interval scale

*Artículo recibido 20 diciembre 2025*

*Aceptado para publicación: 22 enero 2026*



## INTRODUCCIÓN

La metodología tradicional para el desarrollo de instrumentos de medición está basada en la Teoría Clásica de la Medición, y consiste, a grandes rasgos, en llevar a cabo un análisis de ítems, calcular la confiabilidad y determinar la estructura factorial (Kaiser,1974). El análisis de ítems incluye calcular los niveles de dificultad y discriminación, según apliquen, de manera que se optimice la confiabilidad. Entre las limitaciones de este método está el que las escalas que se desarrollan son de tipo ordinal, y no son invariantes. Esto último quiere decir que las puntuaciones de un sujeto en dos instrumentos distintos para medir el mismo constructo no serán iguales.

Una clase de métodos más sofisticados, que se han popularizado debido al aumento en la capacidad de procesamiento de las computadoras, están basados en la Teoría de Respuesta del Ítem (IRT, por sus siglas en inglés). Estos métodos son superiores en cuanto a su precisión y confiabilidad. El método IRT de tres parámetros que se utiliza en el desarrollo de pruebas de logro estandarizadas modernamente, trata de explicar el mayor por ciento posible de la varianza de los datos.

El Modelo de Rasch, que se puede considerar un modelo IRT de un parámetro, tiene como objetivo desarrollar una escala de medición que sea invariante a través de los examinados y someter a prueba la hipótesis de que los datos se ajustan al modelo. El Modelo de Rasch para el desarrollo de pruebas de logro de aprendizajes, cuyos ítems se evalúan de forma dicótoma, ha sido extendido por Andrich (1978) para desarrollar instrumentos de medición de otros constructos, cuyos ítems se evalúan mediante escalas tipo Likert. A diferencia de los instrumentos desarrollados mediante las teorías clásica o IRT, un instrumento desarrollado utilizando modelos de Rasch tienen varias propiedades de interés: es unidimensional, su nivel de medición es de intervalos, y es invariante.

En este escrito se presentan los conceptos básicos de los modelos de Rasch para el desarrollo de instrumentos de medición y se ilustra su aplicación mediante un ejemplo de carácter metodológico. Dicho ejemplo se apoya en una matriz de **datos simulados**, utilizada exclusivamente con fines ilustrativos, sin corresponder a una aplicación empírica ni pretender inferencias sobre una población real. La discusión se concentra en el modelo de Rasch para ítems dicotómicos, aunque se reseñan algunas características del modelo para ítems ordinales, particularmente en el contexto de escalas tipo Likert.



## DESARROLLO

### El concepto de medición

En un artículo clásico, Stevens (1946) definió la medición como una regla para asignar valores numéricos a niveles de una variable. Utilizando esta definición, desarrollo ‘niveles’ de medición, representados por escalas de tipo nominal, ordinal, de intervalo y de razón. Por su parte Rasch (1960/1980) argumentó que la medición es más que la simple asignación de números a niveles de la variable. Para Rasch, debe haber una unidad de medida, como es costumbre en las ciencias naturales (por ejemplo, pulgadas para medir distancias), que se repite para establecer la escala (e.g., 12 pulgadas en una regla). De esta manera se cumple con las expectativas de que haya una correspondencia entre el nivel de la variable medida y las marcas en la escala y de que la medida sea lineal en su relación con el constructo (Wright & Masters, 1982; Wright & Stone, 1979) . Esto último significa que un múltiplo de la unidad de medida debe representar la misma cantidad del constructo a lo largo de la escala, o en términos de Stevens, que la escala sea de intervalos.

Si visualizamos una variable como una línea recta, entonces el proceso de medición consiste en ubicar a los individuos en esa recta de acuerdo con su nivel en la variable. Para que esto tenga sentido práctico, hay que colocar marcas en la recta que representen valores definidos de la variable. La calibración de los ítems en el modelo de Rasch hace esta función. Más aún, el nivel asignado a los sujetos en la variable debe ser independiente de los ítems específicos utilizados en la calibración. Es decir, si hay dos instrumentos utilizados para medir un constructo, y ambos cumplen con los supuestos del modelo de Rasch, el nivel estimado para un individuo deber ser el mismo, dentro de un margen de error razonable, aun cuando los instrumentos sean distintos en cuanto a los ítems que incluyen y del número de estos. De esta manera, el nivel del individuo se mide de forma objetiva y es invariante con respecto al instrumento utilizado. La objetividad en la medición se logra en el modelo de Rasch ya que el modelo matemático se basa solamente en el nivel de dificultad del ítem y en el nivel de habilidad del individuo y no de factores externos tales como el instrumento o la muestra utilizada en la calibración (Andrich, 1978; Andrich, 2004; Andrich & Marais, 2019).



## Modelo matemático

### Modelo ítems dicotómicos

Los ítems dicotómicos son aquellos que se tienen solo dos posibilidades de respuesta. Por ejemplo, en una prueba de aprovechamiento con ítems cierto/falso, la respuesta a la respuesta correcta se le asigna un valor de 1, mientras que a la incorrecta un valor de 0. El modelo de Rasch dicotómico (Andrich, 2005) se basa en la idea de que una persona con mayor habilidad que otra debe tener mayor probabilidad de contestar correctamente un ítem y, de forma similar, si un ítem es más difícil que otro, significa que la probabilidad de contestar correctamente el segundo ítem debe ser mayor que en el primero para cualquier sujeto.

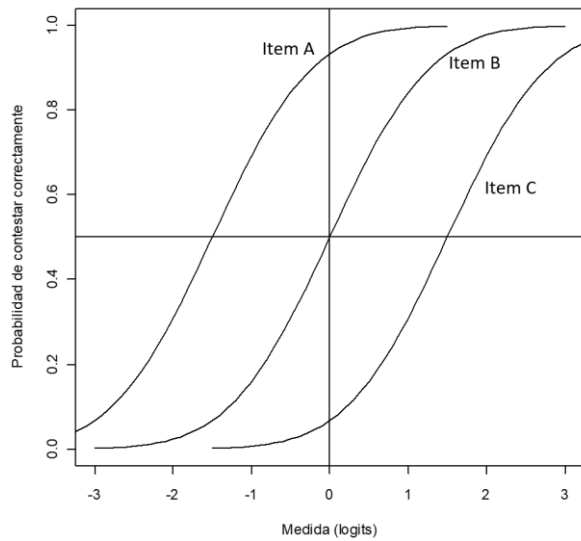
En el modelo de Rasch dicotómico la probabilidad de respuesta de un sujeto a un ítem se modela mediante una función logística del nivel de dificultad y la habilidad del sujeto:

$$P_{ni}\{X_{ni} = 1\} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

En esta expresión matemática,  $P_{ni}$  representa la probabilidad de que la persona  $n$  con habilidad  $\beta_n$  conteste correctamente el ítem con nivel de dificultad  $\delta_i$ . En este sentido el modelo Rasch establece una correspondencia entre la habilidad de los sujetos y los ítems del instrumento (Bond & Fox 2015; Boone, 2020; Boone, Staver & Yale, 2014). La Figura 1 muestra la representación gráfica de la ecuación anterior, llamada ojiva, para tres ítems de diferente nivel de dificultad. En el eje horizontal se representa la medida de la habilidad de los sujetos y en el vertical la probabilidad de contestar correctamente. Como se puede observar, para un sujeto de habilidad media ( $x = 0$ ), la probabilidad de contestar correctamente el ítem C es de alrededor de 0.08, la de contestar correctamente el ítem B es de 0.50, mientras que la de contestar correctamente el ítem A es de alrededor de 0.9. Es decir que el ítem C es más difícil que el ítem B, y este último más difícil que el ítem A.



**Figura 1.** Gráficas de probabilidad acumulada (ojivas) para tres ítems de distinto nivel de dificultad



### Modelo ítems en escala Likert

Se han desarrollado extensiones al modelo de Rasch para considerar instrumentos en los que ítems no son dicótomos. Uno de ellos, desarrollado por Andrich (1978, 1979), permite someter al análisis de Rasch instrumentos en que los ítems individuales se miden mediante escalas de Likert. El modelo

$$\text{matemático en este caso es: } P_{nij} = \frac{\exp(\beta_n - \delta_i - \varphi_j)}{1 + \exp(\beta_n - \delta_i - \varphi_j)}$$

Observe que el modelo es similar al utilizado en el caso de ítems dicótomos, excepto por el factor  $\varphi_j$ . Este parámetro es el que nos permite modelar la categoría de respuestas dentro de la escala Likert. El análisis de Rasch en estos casos es similar al que se presenta en este escrito. Utilizando este modelo se puede evaluar, además, si las respuestas a las categorías de la escala Likert están ordenadas (Wilson, 2005; Embretson & Reise, 2000; Fisher, 2006; Fisher, Stenner, Stone, et al 2021), como normalmente asumimos. Es decir, podemos evaluar de manera empírica si las categorías, digamos 1, 2, 3, 4, 5 corresponden a una mayor cantidad del constructo medido. Cuando esto no ocurre, decimos que las categorías del ítem están desordenadas. Así, por ejemplo, el ordenamiento empírico para las categorías de respuesta de un ítem podría ser en vez 1, 2, 3, 5, 4. En la práctica esto ocurre mayormente, pero no exclusivamente, en escalas Likert con un número de categorías mayor de 5. En tales casos el analista debe decidir si se elimina o se cambia la redacción del ítem. Otra posible solución consiste en colapsar categorías. Es decir, en el caso anterior recodificar las respuestas de manera que 4 y 5 sean una sola

categoría. Las respuestas posibles al ítem serían entonces solo 1, 2, 3, 4.

### El patrón de Guttman

El concepto de Rasch de que un sujeto con mayor habilidad que otro debe tener mayor probabilidad de contestar correctamente un ítem queda ilustrado en el patrón de Guttman (1950). Los datos de una prueba se colocan en una tabla, de manera que las filas correspondan, en orden ascendente de habilidad, a los sujetos, y las columnas, ordenadas en orden ascendente de dificultad, a los ítems. En el cuerpo de la tabla aparecerán ceros y unos, representando las respuestas incorrectas o correctas de los sujetos. Si los datos se ajustan al modelo de Rasch, entonces se podrá observar que tanto los 0, como los 1, toman un patrón triangular. La Tabla 1 presenta un ejemplo del patrón de Guttman idealizado para una escala de diez ítems contestada por diez sujetos.

**Tabla 1.** Patrón de Guttman idealizado para un instrumento de 10 ítems contestado por 10 sujetos

		Ítems									
Sujetos	it05	it02	it03	it06	it04	it07	it10	it08	it01	it09	
P9	1	0	0	0	0	0	0	0	0	0	
P6	1	1	0	0	0	0	0	0	0	0	
P3	1	1	1	0	0	0	0	0	0	0	
P10	1	1	1	1	0	0	0	0	0	0	
P4	1	1	1	1	1	0	0	0	0	0	
P8	1	1	1	1	1	1	0	0	0	0	
P1	1	1	1	1	1	1	1	0	0	0	
P7	1	1	1	1	1	1	1	1	0	0	
P5	1	1	1	1	1	1	1	1	1	0	
P2	1	1	1	1	1	1	1	1	1	1	

Observe, por ejemplo, que el ítem 5, el de menor dificultad, lo contestan correctamente todos los sujetos. El ítem 2 lo contesta incorrectamente solo el sujeto P9 quien tiene el menor nivel de habilidad. Si continuamos hasta el ítem 9, vemos que este, siendo el más difícil de todos, lo contesta correctamente solo el sujeto P2, quien es el de mayor habilidad. En la práctica, el ítem It05 no permite discriminar entre sujetos de alta y baja habilidad, razón por la cual, al igual que ítems que ningún sujeto pueda contestar correctamente, deben ser eliminados del análisis de Rasch.

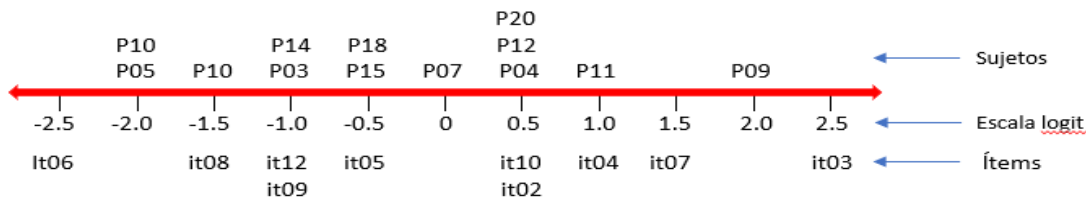
### Mapa Ítem-Sujeto

El resultado del análisis de Rasch es una escala, en unidades logit, que queda definida por la posición de los ítems y en la cual se pueden ubicar los sujetos de acuerdo con su nivel de habilidad (Véase la



Figura 1).

**Figura 1.** Mapa ítem sujetos típico



En esta escala tanto el nivel de dificultad de los ítems como la habilidad de los sujetos aumenta a medida que nos movemos de izquierda a derecha. De esta manera, el sujeto P09 es el que tiene la habilidad más alta, mientras que el ítem it03 es el ítem más difícil. Los ítems it09 e it12 tienen el mismo nivel de dificultad. Uno de estos ítems podría, por lo tanto, eliminarse de la escala, por ser redundante. La escala es lineal de manera que, por ejemplo, la diferencia en habilidad de los sujetos P05 y P03 es la misma que entre los sujetos P11 y P09. Es decir, la escala es de intervalos.

### Calibración de los ítems

Considere un instrumento compuesto por **a** ítems y que es contestado por **n** sujetos. Para cada ítem, **i**, podemos calcular la proporción de sujetos que lo contestan de forma correcta, **pi**, e incorrecta, **1-pi**. Con estas probabilidades se calcula la puntuación logit de incorrectas, mediante la fórmula:

$$x_i = \ln\left(\frac{1 - p_i}{p_i}\right)$$

Estas puntuaciones se multiplican entonces por un factor de corrección para lograr que la ojiva logística se aproxime más a la ojiva normal. Como último paso de la calibración, se le resta a cada puntuación corregida  $x_i$  el promedio de todas las puntuaciones corregidas ( $\bar{x}$ ).

### Medida de la habilidad de los sujetos

El procedimiento para asignar un valor a la habilidad de los sujetos es similar al utilizado para calibrar los ítems. Para cada sujeto, podemos calcular la proporción de ítems contestados correctamente, dividiendo la puntuación obtenida, **s**, entre el número de ítems, **a**:  $p_s = s/a$ . Entonces se calcula la puntuación logit correctos mediante la ecuación:  $y_s = \ln\left(\frac{p_s}{1-p_s}\right)$ . Esta medida inicial se multiplica también por un factor de corrección para obtener el estimado final de la habilidad de cada sujeto.



## **Análisis en el Modelo de Rasch**

Los valores observados para cada sujeto en cada ítem se pueden comparar con los valores esperados según el modelo de Rasch utilizando la fórmula (para el modelo dicotómico):  $P_{ni} = \frac{\exp(B_n - D_i)}{1 + \exp(B_n - D_i)}$ . Donde  $B_n$  representa la habilidad esperada del sujeto y  $D_i$  la dificultad esperada del ítem. La diferencia entre los valores observados y los esperados se pueden utilizar para obtener una suma de cuadrados (Linacre, 2002; Linacre, 2012; Linacre & Smith, 2003; Liu, 2020). Esta última, a su vez, se puede transformar a una estadística que sigue aproximadamente una distribución normal. Esta estadística, denominada 'infit', se puede utilizar para evaluar si un ítem se ajusta al modelo de Rasch. Un ítem con valor del infit entre  $\pm 2.5$ , representa a un ítem que se ajusta al modelo de Rasch. Los ítems que no se ajustan al modelo se pueden analizar para determinar si se pueden revisar para lograr un mejor ajuste, o si hay que descartarlos.

En el modelo de Rasch hay simetría entre los ítems y los sujetos (Wilson & Gochyyev 2020; Willmore Metivier & Santos Abreu, 2025). De esta manera se puede analizar a los sujetos individuales y determinar si se ajustan o no al modelo de Rasch, utilizando también la estadística infit.

## **Análisis de Confiabilidad**

En el modelo de Rasch se pueden calcular estadísticas para determinar la confiabilidad de un instrumento de medición. Por la simetría entre ítems y sujetos descrita arriba, hay dos indicadores de confiabilidad: el índice de confiabilidad de los ítems y el índice de confiabilidad de las personas (Wright & Stone, 1998; Linacre, 2020). El índice de confiabilidad de los ítems cuantifica hasta qué punto el ordenamiento de los ítems se mantendría igual al administrar el instrumento a una muestra de igual tamaño. Este índice es el equivalente al coeficiente de confiabilidad de Cronbach y su magnitud se interpreta de manera similar. El índice de confiabilidad de las personas cuantifica hasta qué punto el ordenamiento de las personas se mantendría igual si utilizáramos un conjunto de ítems distintos.

El índice de separación de las personas es un indicador del poder del constructo medido para discriminar entre los individuos. Su valor mínimo aceptable es de 0.70 e indica que el instrumento es capaz de discriminar entre dos grupos de sujetos (alto y bajo aprovechamiento). Un valor de 0.9, por otro lado, es indicativo de que el instrumento puede discriminar entre cuatro grupos de sujetos.



A continuación, se presenta un **ejemplo de carácter metodológico**, cuyo propósito es ilustrar el procedimiento de análisis desde la perspectiva del modelo de Rasch. El ejemplo se basa en una matriz de respuestas **simuladas**, construida exclusivamente con fines ilustrativos, y no corresponde a una aplicación empírica ni pretende generar inferencias sobre una población real.

En este ejemplo consideraremos una prueba de logro de aprendizajes con 20 ítems, administrada a un grupo de 15 estudiantes. La tabla de datos se ilustra en la Tabla 2. Note que el ítem 11 no fue contestado correctamente por ninguno de los estudiantes. El ítem, por lo tanto, no permite estimar los niveles de habilidad de los estudiantes, razón por la cual se debe eliminar de los análisis. Para este ejemplo se hizo uso del programa para el análisis de Rasch RUMM 2030.

**Tabla 2.** Tabla de datos para el ejemplo del análisis de Rasch dicotómico

	it01	it02	it03	it04	it05	it06	it07	it08	it09	it10	it11	it12	it13	it14	it15	it16	it17	it18	it19	it20
P01	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
P02	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0
P03	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	1	0	1	0
P04	1	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1
P05	0	1	0	1	0	1	1	0	0	0	0	0	0	1	0	1	1	1	1	0
P06	1	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0
P07	1	0	1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	0
P08	0	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1	0	0	0
P09	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0
P10	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
P11	1	1	1	1	0	1	1	0	1	0	0	1	1	0	0	0	1	0	1	0
P12	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
P13	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P14	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	1	0	0
P15	1	1	1	1	0	1	1	0	1	0	0	1	1	0	0	1	1	1	0	0

### Estadísticas de ajuste al modelo

El primer paso en el análisis de Rasch de un instrumento consiste en evaluar si los datos se ajustan al modelo. Esto es similar al análisis factorial confirmatorio, en donde se evalúa también el ajuste de los datos al modelo propuesto. En el análisis de Rasch este ajuste también se puede llevar a cabo mediante la prueba de Chi-cuadrada. Otra estadística global para evaluar el ajuste es el índice de separación de personas. En nuestro ejemplo las estadísticas de resumen que presenta el programa RUMM 2030 muestran que los datos se ajustan al modelo de Rasch (estadística Chi-cuadrada no significativa) con un coeficiente de confiabilidad tradicional (alfa de Cronbach) de 0.60, y los coeficientes de separación de personas igual a 0.65 (Véase la Figura 2). En este caso vemos que los resultados son un tanto ambiguos, pues, aunque la estadística Chi-cuadrada no es significativa, el índice de separación de personas no llega



al mínimo deseado de 0.70. En esta situación es importante evaluar si los ítems individuales se ajustan al modelo, para descartar los que no lo hagan. Otra posibilidad consiste en añadir ítems adicionales al instrumento.

**Figura 2.** Estadísticas de ajuste al modelo de Rasch

ITEM - TRAIT INTERACTION		RELIABILITY INDICES	
Total - Item Chi Square	35.849	PerSepIdx: InitDic2	
Degrees of Freedom	38	* with extms	0.64643
Chi Square Probability	0.569328	* NO extms	0.64643
		CoefficAlpha	
		* with extms	0.60046
		* NO extms	0.60046

El patrón de Guttman para los datos de este ejemplo se ilustra en la Tabla 3. Note que, en términos generales, los ceros se encuentran acumulados hacia la derecha y hacia arriba de la tabla. Las diferencias con el patrón triangular idealizado revelan que los datos, como es de esperarse en la práctica, no se ajustan completamente al modelo de Rasch. Por este motivo es que se llevan a cabo pruebas de significación, como la de Chi-cuadrada mencionada arriba, para determinar si las diferencias con el caso ideal son o no significativas.

Los problemas de ajuste de los ítems y los sujetos al modelo se pueden apreciar fácilmente en el patrón de Guttman. Por ejemplo, un ítem que aparenta no ajustarse al modelo de Rasch es el 12, en el cual a medida que aumenta la habilidad de los sujetos, las respuestas se alternan entre ceros y unos. Los ítems 10 y 19, por otro lado, se ajustan bien al modelo puesto que solo los estudiantes con el más alto nivel de habilidad, P15 y P04 los contestan correctamente.

Desde el punto de vista de los estudiantes, podemos ver que P15 tiene un buen ajuste al modelo. Este estudiante fue capaz de contestar correctamente los doce ítems más fáciles, mientras que contestó incorrectamente los difíciles. El estudiante P03 parece un caso extremo de desajuste, pues es capaz de contestar alternadamente los ítems según aumenta el nivel de dificultad de estos. Una posible explicación en este caso es que el estudiante contestó los ítems al azar. En términos de calibrar los ítems del instrumento, es recomendable eliminar los sujetos cuyas respuestas se alejan mucho del comportamiento esperado y rehacer los análisis.

Debe ser obvio para el lector que este análisis cualitativo se dificulta cuando el tamaño de la muestra y el número de ítems es grande. En tales casos es mejor guiarse primero por los resultados de la estadística



de ajuste infit y luego, si se desea, examinar el patrón de respuesta de los sujetos cuyo comportamiento es anómalo.

**Tabla 3.** Patrón de Guttman para los datos del ejemplo

	IT04	IT03	IT06	IT16	IT02	IT01	IT12	IT17	IT11	IT15	IT07	IT09	IT18	IT14	IT08	IT13	IT05	IT10	IT19
P01	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P13	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
P12	1	1	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0
P10	1	1	1	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0
P09	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
P14	1	1	1	1	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
P08	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0
P06	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
P07	1	1	1	1	0	1	1	0	1	0	0	0	0	1	0	1	0	0	0
P05	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0
P02	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0
P11	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	0	0	0
P03	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0
P15	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
P04	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	1	1

### Análisis de ítems

La Tabla 4 muestra, para cada ítem, su puntuación en la escala logit (Location), el error estándar asociado a esta medida, y la estadística infit (FitResid). Los valores del infit se encuentran todos dentro del rango  $\pm 2.5$  lo que indica que los ítems individuales tienen un buen ajuste al modelo de Rasch.

**Tabla 4.** Puntuaciones logit (Location), error estándar y estadística de ajuste para los ítems del ejemplo.

Item	Location	SE	FitResid
it01	-1.07	0.60	-0.39
it02	-2.06	0.72	-0.26
it03	-2.88	0.90	0.29
it04	-3.22	1.01	-0.03
it05	2.61	1.03	-0.27
it06	-2.73	0.86	-0.30
it07	0.49	0.59	-0.30
it08	1.42	0.70	1.28
it09	0.67	0.60	0.36
it10	2.81	1.12	-0.49
it11	0.31	0.58	0.65
it12	-0.67	0.58	1.32
it13	1.63	0.74	0.03
it14	0.76	0.61	0.27
it15	0.41	0.58	-0.59
it16	-2.24	0.76	-0.66
it17	0.20	0.57	-0.11
it18	0.76	0.61	-0.69
it19	2.81	1.12	-0.49



## Análisis de los sujetos

La Tabla 5 presenta las estadísticas de los sujetos: su puntuación en la prueba, la medida en la escala logit, el error estándar de la medida y la estadística infit. Esta última muestra que las respuestas de todos los sujetos se ajustan al modelo Rasch. La columna SE muestra una característica adicional del análisis de Rasch: la precisión de la medida es, en general, distinta para cada sujeto. Esto se debe a la ubicación del sujeto en relación con la ubicación de los ítems en la escala logit. Los sujetos cuya medida queda cercana a varios ítems tendrá más precisión que la de los sujetos en los cuales la posición de su medida se encuentre relativamente lejos de los ítems. Esta es una diferencia importante con la teoría clásica de la medición en donde se le asigna un mismo error estándar a todas las medidas.

**Tabla 5.** Puntuación cruda, puntuación logit, error estándar y estadística de ajuste para los sujetos del ejemplo.

Person	Score	Location	SE	FitResid
P01	3	-2.63	0.75	-0.22
P02	9	-0.06	0.60	-1.15
P03	11	0.61	0.59	0.38
P04	13	1.29	0.60	0.14
P05	9	-0.06	0.60	1.19
P06	8	-0.41	0.61	-1.02
P07	9	-0.06	0.60	0.35
P08	8	-0.41	0.61	-0.60
P09	7	-0.79	0.63	0.16
P10	7	-0.79	0.63	-0.18
P11	11	0.61	0.59	-0.80
P12	7	-0.79	0.63	-0.08
P13	4	-2.12	0.71	0.60
P14	8	-0.41	0.61	-0.29
P15	12	0.95	0.59	-0.94

## Mapa ítem-sujeto

El mapa ítem-sujeto para este ejemplo se ilustra en la Figura 3. A la izquierda aparece la escala logit y a la izquierda y la derecha de la línea entrecortada vemos la posición de los sujetos y los ítems en dicha escala. Como se puede apreciar el sujeto con mayor habilidad es P04 y el de menor P01. De igual manera los ítems IT19 e IT10 son los más difíciles, mientras que el ítem IT04 es el más fácil. El mapa ayuda a entender por qué los ítems 3, 11 y 15 son los que tienen el error estándar menor (mayor precisión): en sus cercanías hay ubicados siete ítems (IT09, IT18, IT14, IT15, IT07, IT17 E IT11). El mapa permite



además evaluar cómo mejorar el instrumento. En este caso particular, el instrumento mejoraría si se incluyen ítems con niveles de dificultad intermedios a la de los ítems IT05 e IT13, IT17 e IT09, e IT01 y IT02.

LOCATION	PERSONS	ITEMS
3.0		IT19 IT10 IT05
2.0		IT13 IT08
1.0	P04	
	P15	
	P03 P11	IT09 IT18 IT14 IT15 IT07 IT17 IT11
0.0	P02 P05 P07	
	P06 P08 P14	
	P09 P10 P12	IT12
-1.0		IT01
-2.0	P13	IT02 IT16
-3.0	P01	IT06 IT03 IT04
-4.0		

## CONCLUSIONES

El desarrollo de instrumentos de medición utilizando el modelo de Rasch ofrece varias ventajas sobre la teoría clásica de la medición (Muñiz, 2003). Primero, la escala de medición para los constructos es de intervalos (Wright, & Masters, 1982). En la teoría clásica las escalas son, realmente, ordinales, pero en la práctica son tratadas como si fueran de intervalos, creando inconsistencias entre el análisis paramétrico o no paramétrico de estas variables. Segundo, las medidas son invariantes, es decir, distintos instrumentos (Messick, 1995/1998; Frisbie, 1988; AERA, 2014; Holland & Thayer, 1988 ; Traub & Rowley 1991) para medir un constructo asignarán valores similares a los sujetos. Incluso, con los ítems debidamente calibrados, las puntuaciones logit que asignen los instrumentos serán similares, no importa el número de ítems en el instrumento. Tercero, los instrumentos desarrollados utilizando el modelo de



Rasch son unidimensionales, es decir, consisten en un solo factor, lo que facilita su interpretación.

El análisis de datos en el modelo de Rasch incluye el cálculo de estadísticas de ajuste para los ítems y las personas. De esta manera se puede evaluar si los ítems o las personas individuales se conforman al modelo. La estadística Chi-cuadrada puede ser utilizada para hacer una evaluación global del ajuste de los datos al modelo de Rasch, mientras que el patrón de Guttman permite, cuando el número de ítems y el tamaño de la muestra es pequeño, hacer una evaluación visual del ajuste.

Finalmente, el mapa ítem-sujeto muestra la escala logit y la ubicación en esta escala de los ítems y los sujetos. Por ser una medida lineal se hace fácil convertir esta escala a una con un rango de valores más familiar, como, por ejemplo, entre 0-100.

## REFERENCIAS BIBLIOGRÁFICAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social, and health sciences*. Springer Nature Singapore.

Andrich, D. (2004). *Controversy and the Rasch model: A characteristic of incompatible paradigms?* *Medical Care*, 42(1 Suppl), I-7–I-16. <https://doi.org/10.1097/01.mlr.0000103528.48582.7c>

Andrich, D. (2005). Rasch models. En K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 395–402). Elsevier.

Bond, T. G. y Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Routledge.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge.

Boone, W. J. (2020). *Rasch analysis for instrument development: Why, when, and how?* Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Cano, J., Melin, J., Pendrill, L., Stenner, A. J., Fisher, W. P., & Stenner, P. (2016).



- Developing a common language for measuring human and social capital. *Measurement: Journal of the International Measurement Confederation*, 92, 489–496.  
<https://doi.org/10.1016/j.measurement.2016.06.052>. Commons, M. L., & Goodheart, E. A. (2008). *The systemic and metasystemic stages of performance*. Springer.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engelhard, G. (2013). *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. Routledge.
- Fisher, W. P. (2006). Meaningfulness, measurement, and metrological networks: Standard units as evolving integrating factors. *Measurement*, 39(7), 674–679.  
<https://doi.org/10.1016/j.measurement.2006.05.001>
- Fisher, W. P., Stenner, A. J., Stone, M., & others. (2021). Construct maps and the Wright map revisited. *Journal of Applied Measurement*, 22(2), 105–127.
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *National Council on Measurement in Education*.
- Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer et al. (Eds.), *Measurement and prediction* (pp. 60–90). Princeton University Press.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Lawrence Erlbaum.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–36.  
<https://doi.org/10.1007/BF02291575>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Linacre, J. M. (2012). *Winsteps® Rasch measurement computer program user's guide*. Winsteps.com.
- Linacre, J. M., & Smith, E. V. (2003). A proposal for standardization of Rasch measurement. *Rasch Measurement Transactions*, 17(2), 918–919.
- Liu, X. (2020). *Using and developing measurement instruments in science education: A Rasch modeling approach* (2nd ed.). Information Age Publishing.



- Masters, G. N. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement*, 25(1), 15–29.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Muñiz, J. (2003). *Teoría clásica de los test*. Madrid: Pirámide.
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Sage.
- 167
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV* (pp. 321–334). Berkeley: 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.
- Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests.(Copenhagen, Danish Institute for Educational Research).
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology, pp. 321–334 in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV*. Berkeley: University of Chicago Press, 1980.
- 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.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Traub, R. E., & Rowley, G. L. (1991). *Understanding reliability: An instructional module*. National Council on Measurement in Education (NCME).



- Willmore Metivier, J., & Santos Abreu, C. M. (2025). Análisis de Rasch en la medición de competencias matemáticas en estudiantes de nuevo ingreso de la PUCMM. Comunicación presentada en el IV Congreso CEMACYC, Santo Domingo, República Dominicana <https://ponencias.ciaem-redumate.org/cemacyc/article/view/487/531>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.
- Wilson, M., & Gochyyev, P. (2020). Measurement as a social practice: Constructing measures with construct modeling. *Measurement: Interdisciplinary Research and Perspectives*, 18(1), 1–33. <https://doi.org/10.1080/15366367.2020.1711592>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.
- Wright, B. D., & Stone, M. H. (1998). *Diseño de mejores pruebas utilizando la técnica de Rasch*. México: CENEVAL.
- Wright, B. D., & Stone, M. H. (1999). *Measurement essentials (2nd ed.)*. Wide Range.

